Seminorms for multiple averages along polynomials and applications to joint ergodicity

Preprint · February 2019

Sebastián Donoso
University of Chile

Andreas Koutsogiannis
Aristotle University of Thessaloniki

SEMINORMS FOR MULTIPLE AVERAGES ALONG POLYNOMIALS AND APPLICATIONS TO JOINT ERGODICITY

SEBASTIÁN DONOSO, ANDREAS KOUTSOGIANNIS AND WENBO SUN

Abstract. Exploiting the recent work of Tao and Ziegler on the concatenation theorem on factors, we find explicit characteristic factors for multiple averages along polynomials on systems with commuting transformations, and use them to study the criteria of joint ergodicity for sequences of the form $(T_1^{p_{1,j}(n)}\cdot\ldots\cdot T_d^{p_{d,j}(n)})_{n\in\mathbb{Z}}$, $1\leq j\leq k$, where $T_1,\ldots,T_d$ are commuting measure preserving transformations on a probability measure space and $p_{i,j}$ are integer polynomials. To be more precise, we provide a sufficient condition for such sequences to be jointly ergodic. We also give a characterization for sequences of the form $(T_i^{p(n)})_{n\in\mathbb{Z}}, 1\leq i\leq d$ to be jointly ergodic, answering a question due to Bergelson.

1. INTRODUCTION

1.1. **Characteristic factors for multiple averages.** Let $\mathbf{X}=(X,\mathcal{B},\mu,T)$ be a measure preserving $\mathbb{Z}$-system.[1] When T is ergodic (*i.e.*, the measure of any T-invariant set is 0 or 1), the von Neumann ergodic theorem (see for example [10, Theorem 2.21]) asserts that for all $f\in L^2(\mu)$, the $L^2(\mu)$ limit of the "time average" $\dfrac{1}{N}\sum_{n=0}^{N-1}T^nf$ equals to the "natural" one, namely the "space limit" $\int_X f\,d\mu$.

In the past decades, the L^2-limit behavior of the "multiple averages" became a central topic in ergodic theory. Several authors have studied averages for a single transformation T, as

$$\frac{1}{N}\sum_{n=0}^{N-1}T^{p_1(n)}f_1\cdot\ldots\cdot T^{p_k(n)}f_k, \tag{1}$$

averages for several (usually commuting) T_i's, as

$$\frac{1}{N}\sum_{n=0}^{N-1}T_1^{p_1(n)}f_1\cdot\ldots\cdot T_k^{p_k(n)}f_k \tag{2}$$

and even more general averages as

2010 *Mathematics Subject Classification.* Primary: 37A05; Secondary: 37A30, 28A99.

Key words and phrases. Multiple averages, polynomials, characteristic factors, concatenation theorem.

The first author is supported by Fondecyt Iniciación en Investigación Grant 11160061.

[1] By this we mean that $(X,\mathcal{B},\mu)$ is a probability space and T is an invertible measure preserving transformation, *i.e.*, $\mu(T^{-1}A)=\mu(A)$ for all $A\in\mathcal{B}$. We also denote such a system as $(X,\mathcal{B},\mu,(S_g)_{g\in\mathbb{Z}})$ later in this paper, where $S_n=T^n$, *i.e.*, the composition of T with itself n times.

$$\frac{1}{N}\sum_{n=0}^{N-1}\prod_{i=1}^{m}T_i^{p_{i,1}(n)}f_1\cdot\ldots\cdot\prod_{i=1}^{m}T_i^{p_{i,k}(n)}f_k \tag{3}$$

for some $m, k \in \mathbb{N}^*$, integer valued sequences $(p_i(n))_{n\in\mathbb{N}}$ and $f_i \in L^\infty(\mu)$, $1 \leq i \leq k$. Fruitful results has been obtained, which include, but are not limited to [1, 4, 6, 8, 9, 14, 15, 16, 17, 18, 19, 21, 23, 25, 27]. In particular, it was proved by Walsh [25] (following the ideas of Tao [23]) that the multiple (uniform) averages, as in (3), converge in the L^2 sense for any integer valued polynomials p_i when $T_1, \ldots, T_d$ span a nilpotent group. However, the result in [25] does not give any description or information about the limit. In general, very little is known about the limit of multiple averages.

The existing results employ the idea of *characteristic factors*, which intends to reduce the average under study to a more tractable one. For a single transformation T and for linear p_i's, the main content of [16] is the introduction of some seminorms that control the behavior of the average (1) and are characterized by *nilsystems*. These seminorms were also used by Leibman (in [20]) to bound the limit of (1) for polynomial p_i's (always in the context of a single transformation). For several commuting transformations, Host (in [15]) introduced similar seminorms to bound the limit of (2) for linear p_i's but in that case there was still no clear connection to nilsystems (see also [22, 24] for slight generalizations of these seminorms). When considering non linear polynomials p_i's, even less is known and even simple cases can be very intricate. For instance, Austin in [2, 3] found precise characteristic factors for some specific cases of quadratic polynomials for $k = 2$ (and linear polynomials for $k = 3$).

In this paper, under a further development of a recent result by Tao and Ziegler ([24]) on concatenation (intersection) of factors, we provide an upper bound for the limit of (3) for any $m, k \in \mathbb{N}^*$ and polynomials $p_{i,j}$ taking integer values at integers by using some seminorms on the system (generically called *Host-Kra seminorms*), which to the best of our knowledge, has never been studied before in this generality. We state here a simplified more aesthetic one-parameter version of our main result, and refer the readers to Theorem 5.1 below for the result in its full generality:

Theorem 1.1 (Bounding multiple averages along polynomials by seminorms)**.** *Let $d, k, K \in \mathbb{N}^*$ and $p_1, \ldots, p_k \colon \mathbb{Z} \to \mathbb{Z}^d$ be a family of polynomials of degrees at most K such that $p_i, p_i - p_j$ are not constant for all $1 \leq i, j \leq k$, $i \neq j$, where $p_i(n) = \sum_{0\leq v\leq K} b_{i,v}n^v$ for some $b_{i,v} \in \mathbb{Q}^d$. Denote the set of the coefficients and pairwise differences of the coefficients (excluding $\mathbf{0}$) of the polynomials with*

$$R = \bigcup_{0<v\leq K}\{b_{i,v}, b_{i,v} - b_{i',v} \colon 1 \leq i, i' \leq k\}\backslash\{\mathbf{0}\}.$$

Let $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system (see Section 1.4 for the definition). If the Host-Kra seminorm $\|f_i\|_{\{G(r)^{\times\infty}\}_{r\in R}}$ (see Section 2 for definitions) of f_i equals to 0 for some $1 \leq i \leq k$, then

$$\lim_{N-M\to\infty}\frac{1}{N-M}\sum_{n=M}^{N-1}T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k = 0.$$

Remark. Unlike the conventional "finite-step" Host-Kra seminorms, the seminorms we use such as $\|\cdot\|_{\{G(r)\times\infty\}_{r\in R}}$ are "infinite-step" ones. It is an interesting question to ask whether one can replace the "infinite-step" seminorms in the main theorems of this paper by "finite-step" ones.

1.2. **The joint ergodicity property.** An interesting application of Theorem 1.1 and its stronger version Theorem 5.1 is that they can be used to study joint ergodicity problems, also allowing us to answer a question due to Bergelson. Back to the description of the limit of (3), there are interesting cases where the limit has a "simple" description. In [6], Bergelson showed that if $(X,\mathcal{B},\mu,T)$ is a weakly mixing system (meaning that $T\times T$ is ergodic for $\mu\times\mu$) [2] and $p_1,\dots,p_k$ are polynomials such that p_i, p_i-p_j are non-constant for all $1\leq i,j\leq k, i\neq j$, then the $L^2(\mu)$ limit of (1) is the "expected" one, namely the "multiple space limit" $\prod_{i=1}^{k}\int_X f_i\,d\mu$.[3] One can think of this result as a strong independence property of the sequences $(T^{p_i(n)})_{n\in\mathbb{Z}}, 1\leq i\leq k$ in the weakly mixing case. This naturally leads to the following definition of joint ergodicity, in which we demand the average to converge to the expected limit.

Definition. Let $d,k,L\in\mathbb{N}^*$, $p_1,\dots,p_k\colon\mathbb{Z}^L\to\mathbb{Z}^d$ be functions, and $(X,\mathcal{B},\mu,(T_g)_{g\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system. We say that the tuple $(T_{p_1(n)},\dots,T_{p_k(n)})_{n\in\mathbb{Z}^L}$ is *jointly ergodic* for μ if for every $f_1,\dots,f_k\in L^\infty(\mu)$ and every Følner sequence $(I_N)_{N\in\mathbb{N}}$ of $\mathbb{Z}^L$,[4] we have that

$$\lim_{N\to\infty}\frac{1}{|I_N|}\sum_{n\in I_N}T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k=\int_X f_1\,d\mu\cdot\ldots\cdot\int_X f_k\,d\mu, \tag{4}$$

where the limit is taken in $L^2(\mu)$. When $k=1$, we also say that $(T_{p_1(n)})_{n\in\mathbb{Z}^L}$ is *ergodic* for μ instead.[5]

For $d,L\in\mathbb{N}^*$, we say that $q\colon\mathbb{Z}^L\to\mathbb{Z}^d$ is an *integer-valued polynomial* if $q=(q_1,\dots,q_d)$, where each q_i is an integer polynomial (meaning that it takes integer values at integers) of L variables. The polynomial q is *non-constant* if some q_i is non-constant. A family of polynomials $p_1,\dots,p_k\colon\mathbb{Z}^L\to\mathbb{Z}^d$ is *non-degenerate* if they are *essentially non-constant* (meaning that each p_i is not a constant polynomial) and *essentially distinct* (meaning that p_i-p_j is essentially non-constant for all $1\leq i,j\leq k, i\neq j$).[6] Using this new language, it follows from [6] that if T is weakly mixing and $p_1,\dots,p_k\colon\mathbb{Z}\to\mathbb{Z}$ is a non-degenerate family of polynomials, then $(T^{p_1(n)},\dots,T^{p_k(n)})_{n\in\mathbb{Z}}$ is jointly ergodic for μ. Later, it was proved by Frantzikinakis and Kra (in [14]) that if $p_1,\dots,p_k\colon\mathbb{Z}\to\mathbb{Z}$ is an independent family of polynomials (*i.e.*, every linear combination along integers of the p_i's is non-constant) and T is *totally ergodic* (*i.e.*, T^n is ergodic

[2] In this case we also say that T is a weakly mixing transformation.

[3] This result was previously obtained by Furstenberg (in [11]) in the special case where $p_i(n)=in$, $i=1,\dots,k$.

[4] A sequence of finite subsets $(I_N)_{N\in\mathbb{N}}$ of $\mathbb{Z}^L$ with the property $\lim_{N\to\infty}|I_N|^{-1}\cdot|(g+I_N)\triangle I_N|=0$ for all $g\in\mathbb{Z}^L$, is called *Følner sequence* in $\mathbb{Z}^L$.

[5] The main reason we change from single-variable p_i's to multi-variable ones and give the definition in this generality is technical. More specifically, we will deal with multi-variable integer valued polynomials, since our arguments, even for single-variable polynomials, naturally lead to multi-variable ones (for details, see the "dimension-increment" method, explained before Proposition 6.3).

[6] Throughout this paper, when we write "a polynomial $p\colon\mathbb{Z}^L\to\mathbb{Z}^d$", we implicitly assume that p is integer-valued, hence, in general, p has rational coefficients.

for all $n \in \mathbb{Z}\backslash\{0\}$), then the tuple $(T^{p_1(n)}, \dots, T^{p_k(n)})_{n\in\mathbb{Z}}$ is jointly ergodic for μ (for integer part of real valued strongly independent polynomials, see [18]). By combining existing results, we have the following proposition:

Proposition 1.2. *Let $d, k, L \in \mathbb{N}^*$ and $p_1, \dots, p_k\colon \mathbb{Z}^L \to \mathbb{Z}^d$ be a non-degenerate family of polynomials. Let $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system such that:*

(i) *T_g is ergodic for μ for all $g \in \mathbb{Z}^d \setminus \{\mathbf{0}\}$; and*
(ii) *$(T_{p_1(n)} \times \cdots \times T_{p_k(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes k}$.*[7]

Then $(T^{p_1(n)}, \dots, T^{p_k(n)})_{n\in\mathbb{Z}^L}$ is jointly ergodic for μ.

Proposition 1.2 is a direct corollary of Proposition 2.10 of [17], Theorem 10.1 of [16] (see also Theorem 2.6 below), and Theorem B of [20] (see also Theorem 2.9 below). We leave the details of the proof to the interested readers.

We remark that in all the aforementioned results, one needs to make rather strong assumptions for the system, more specifically that either the transformation is weakly mixing or that infinitely many transformations T_g are ergodic. It is then natural to ask if one can obtain joint ergodicity results under weaker conditions, *e.g.*, assuming that only finitely many transformations (or sequences of transformations with specific iterates) are ergodic, and finally, if there are any cases in which the sufficient condition is also necessary. In this direction, it is worth mentioning two results related to our study.

Let $d \in \mathbb{N}^*$ and $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a measure preserving system with commuting transformations.[8] It was proved by Berend and Bergelson (in [4]) that the tuple $(T_1^n, \dots, T_d^n)_{n\in\mathbb{Z}}$ is jointly ergodic for μ if and only if $T_iT_j^{-1}$ is ergodic for μ for all $1 \leq i, j \leq d, i \neq j$ and $T_1 \times \cdots \times T_d$ is ergodic for $\mu^{\otimes d}$. Recently, it was proved by Bergelson, Leibman and Son (in [8]) that if $p_1, \dots, p_d\colon \mathbb{Z} \to \mathbb{Z}$ are generalized linear functions (*i.e.*, functions of the form $p(n) = [\alpha_1 n + \alpha_2]$, $[\alpha_3[\alpha_1 n + \alpha_2]]$, *etc.*, where $[\cdot]$ denotes the integer part, or floor, function), then the tuple $(T_1^{p_1(n)}, \dots, T_d^{p_d(n)})_{n\in\mathbb{Z}}$ is jointly ergodic for μ if and only if the sequence $(T_i^{p_i(n)}T_j^{-p_j(n)})_{n\in\mathbb{Z}}$ is ergodic for μ for all $1 \leq i, j \leq d, i \neq j$, and the sequence $(T_1^{p_1(n)} \times \cdots \times T_d^{p_d(n)})_{n\in\mathbb{Z}}$ is ergodic for $\mu^{\otimes d}$. Note that both results, while being characterizations, hold under only the ergodicity assumption for finitely many transformations and sequences of transformations.

In this paper, we study joint ergodicity properties for sequences of transformations with polynomial iterates. The following is our first application of Theorems 1.1 and 5.1:

Theorem 1.3. *Let $d, k, K, L \in \mathbb{N}^*$ and $p_1, \dots, p_k\colon \mathbb{Z}^L \to \mathbb{Z}^d$ be a non-degenerate family of polynomials of degrees at most K. Suppose that $p_i(n) = \sum_{v\in\mathbb{N}^L, |v|\leq K} b_{i,v} n^v$ for some $b_{i,v} \in \mathbb{Q}^d$.*[9] *Denote the set of the coefficients and pairwise differences of the coefficients (excluding $\mathbf{0}$) of the polynomials with*

$$R = \bigcup_{0<|v|\leq K} \{b_{i,v}, b_{i,v} - b_{i',v}\colon 1 \leq i, i' \leq k\}\backslash\{\mathbf{0}\}.$$

[7] $\mu^{\otimes k}$ is the product measure $\mu \otimes \cdots \otimes \mu$ on X^k.

[8] Here, as in the expression (2), $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ can be understood as an abbreviation for the $\mathbb{Z}^d$-system $(X, \mathcal{B}, \mu, (S_g)_{g\in\mathbb{Z}^d})$, where $T_1 = S_{(1,0,\dots,0)}, T_2 = S_{(0,1,0,\dots,0)}, \dots, T_d = S_{(0,\dots,0,1)}$.

[9] For $n = (n_1, \dots, n_L) \in \mathbb{Z}^L$ and $v = (v_1, \dots, v_L) \in \mathbb{N}^L$, n^v denotes the quantity $n_1^{v_1} \cdot \ldots \cdot n_L^{v_L}$, and $|v| = v_1 + \cdots + v_L$.

Let $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system such that:

(i) *For all $r \in R$, denoting $G(r) := span_{\mathbb{Q}}\{r\} \cap \mathbb{Z}^d$ (see also the relation (7) in the corresponding definition in Subsection 2.5), the action $(T_g)_{g\in G(r)}$ is ergodic for μ;*[10] *and*

(ii) *$(T_{p_1(n)} \times \cdots \times T_{p_k(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes k}$.*

Then $(T_{p_1(n)}, \dots, T_{p_k(n)})_{n\in\mathbb{Z}^L}$ is jointly ergodic for μ.

We remark that Theorem 1.3 is stronger than Proposition 1.2 since we only require finitely many T_g's to be ergodic, *i.e.*, those g's belonging to R, and the set R has an explicit expression.

Example 1. Let $(X, \mathcal{B}, \mu, T_1, T_2)$ be a system with two commuting transformations and assume that $(T_1^{n^2+n} \times T_2^{n^2})_{n\in\mathbb{Z}}$ is ergodic for $\mu \times \mu$. Then Theorem 1.3 implies that if $T_1, T_2, T_1T_2^{-1}$ are ergodic for μ, then $(T_1^{n^2+n}, T_2^{n^2})_{n\in\mathbb{Z}}$ is jointly ergodic for μ.

Conversely, the joint ergodicity of $(T_1^{n^2+n}, T_2^{n^2})_{n\in\mathbb{Z}}$ implies the ergodicity of $(T_1^{n^2+n})_{n\in\mathbb{Z}}$ and $(T_2^{n^2})_{n\in\mathbb{Z}}$ for μ, which in turn implies the ergodicity of T_1 and T_2 for μ. However, the fact that $(T_1^{n^2+n}, T_2^{n^2})_{n\in\mathbb{Z}}$ is jointly ergodic for μ does not necessarily imply that $T_1T_2^{-1}$ is ergodic (take for instance $T_1 = T_2 = T$ where T is a weakly mixing transformation).

Throughout this paper, Example 1 will be our main example via which we demonstrate how our method works. Note that annoyingly enough, the expression of the limit of the average of the sequence $T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2$ for bounded f_1 and f_2 cannot be immediately found from known results, despite the fact that the polynomials $p_1(n) = n^2 + n$ and $p_2(n) = n^2$ are essentially distinct.

The second application of Theorems 1.1 and 5.1 is the following theorem, which provides necessary and sufficient conditions for joint ergodicity of the polynomial sequences $T_i^{p(n)}, 1 \leq i \leq d$. This generalizes the result from [4] and answers a question due to Bergelson:[11]

Theorem 1.4. *Let $d, L \in \mathbb{N}^*$, $p\colon \mathbb{Z}^L \to \mathbb{Z}$ be a polynomial and $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a system with commuting transformations. Then $(T_1^{p(n)}, \dots, T_d^{p(n)})_{n\in\mathbb{Z}^L}$ is jointly ergodic for μ if and only if both of the following conditions are satisfied:*

(i) *$T_iT_j^{-1}$ is ergodic for μ for all $1 \leq i, j \leq d$, $i \neq j$; and*

(ii) *$((T_1 \times \cdots \times T_d)^{p(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes d}$.*

As an immediate example, for a system $(X, \mathcal{B}, \mu, T_1, T_2)$ with two commuting transformations, the sequence $(T_1^{n^2}, T_2^{n^2})_{n\in\mathbb{Z}}$ is jointly ergodic for μ if and only if $T_1T_2^{-1}$ is ergodic for μ and $(T_1^{n^2} \times T_2^{n^2})_{n\in\mathbb{Z}}$ is ergodic for $\mu \times \mu$.

One might wonder if there are better descriptions of condition (ii) of Theorem 1.4. In Section 3, we provide several criteria and equivalent conditions of (ii), related to the eigenvalues of the system.

Based on the work of [4, 8] and the main results of this paper, we have a natural conjecture:

---

[10] For a subgroup H of $\mathbb{Z}^d$, $(T_g)_{g\in G(r)}$ is *ergodic* for μ if every $A \in \mathcal{B}$ which is invariant under T_g for all $g \in G(r)$ is of μ-measure 0 or 1.

[11] Personal communication.

Conjecture 1.5. *Let $d, k, L \in \mathbb{N}^*$, $p_1, \ldots, p_k \colon \mathbb{Z}^L \to \mathbb{Z}^d$ be polynomials and $(X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system. Then $(T_{p_1(n)}, \ldots, T_{p_k(n)})_{n \in \mathbb{Z}^L}$ is jointly ergodic for μ if and only if both of the following conditions are satisfied:*

(i) *$(T_{p_i(n)-p_j(n)})_{n \in \mathbb{Z}^L}$ is ergodic for μ for all $1 \leq i, j \leq k$, $i \neq j$; and*
(ii) *$(T_{p_1(n)} \times \cdots \times T_{p_k(n)})_{n \in \mathbb{Z}^L}$ is ergodic for $\mu^{\otimes k}$.*

1.3. **Method and Organization.** Section 2 contains all the background material and Section 3 the conditions equivalent to (ii) of Theorem 1.4 (see Proposition 3.2).

In order to prove the joint ergodicity results of this paper, we introduce a characterization theorem (Theorem 5.1, the stronger version of Theorem 1.1) in Section 5, which allows us to study joint ergodicity properties under the assumption that all the functions $f_1, \ldots, f_k$ are measurable with respect to certain Host-Kra characteristic factors (see Section 2 for definitions). Once Theorem 5.1 is proven, a standard argument using results from [16, 21] (or via Theorems 2.6 and 2.9 – see below) yields the main results of this paper. The proofs of Theorems 1.3 and 1.4, under the assumption of the validity of Theorem 5.1, are enclosed in Section 5 as well. In the same section, we also introduce the two main ingredients for proving Theorem 5.1, namely Propositions 5.5 (which we prove in Section 6) and 5.6 (which we prove in Section 7).

To obtain the characterization theorem (Theorem 5.1), we employ the by now classical "PET induction" (first introduced in [6]), which allows us to convert the average in (4) to a special case where every $p_i(n)$ is a linear function by repeatedly applying the van der Corput lemma (Lemma 2.2). Adaptations of this method have been extensively studied in the past in [9, 17, 21] too. We explain it in detail in Section 4 tailored for our purposes.

There are two major difficulties to carry out the PET induction in proving Theorem 5.1 though. The first is that although PET induction variations used in the past allow us to eventually reduce the left hand side of (4) to an expression with linear iterates, they provide no information on the coefficients of these iterates, which is a crucial detail in describing the set R defined in Theorem 1.3. To overcome this difficulty, we introduce a new alteration of this technique in Section 6 (see the proof of Proposition 5.5) which allows us to keep track of the coefficients of the polynomials when we iteratively apply van der Corput (vdC) operations.

The second, and perhaps the most important problem is how to bound the left hand side of (4) by some Host-Kra-type seminorm of each function f_i. It turns out that for a general non-degenerate family of polynomials $p_1, \ldots, p_k \colon \mathbb{Z}^L \to \mathbb{Z}^d$, we can use the PET induction to bound the left hand side of (4) by an averaged Host-Kra seminorm, as the righthand sides of (21) and (23) (see Section 5). The problem-goal now is to bound such an averaged seminorm effectively by a single one. In the past, in analogous situations, issues like these were resolved under additional restrictions, such as the assumption that $d = 1$ ([6]), that all T_g's are ergodic ([14, 17]), or that $p_1, \ldots, p_k$ have different (and positive) degrees ([9]). In this paper, we address this difficulty in Section 7 (see the proof of Proposition 5.6) in its full generality. Our method is based on the recent work of Tao and Ziegler on the concatenation theorem ([24]).

Acknowledgements. We thank Vitaly Bergelson for bringing the problem that we are addressing in Theorem 1.4 to our attention, and also for providing useful advices. The second author thanks the hospitality of the Center for Mathematical Modeling (CMM) of the University of Chile, where this work started. The first author thanks the hospitality of the Ohio State University, where this work was finished.

1.4. **Definitions and notations.** We denote with $\mathbb{N}^*$, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ the set of positive integers, non-negative integers, integers, rational numbers and real numbers, respectively. If X is a set, and $d \in \mathbb{N}^*$, X^d denotes the Cartesian product $X \times \cdots \times X$ of d copies of X.

We say that a tuple $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ is a $\mathbb{Z}^d$-*measure preserving system* (or a $\mathbb{Z}^d$-*system*) if $(X, \mathcal{B}, \mu)$ is a probability space and $T_g \colon X \to X$ are measurable, measure preserving transformations on X such that $T_{(0,\dots,0)} = id$ and $T_g \circ T_h = T_{gh}$ for all $g, h \in \mathbb{Z}^d$. The system is *ergodic* if for any $A \in \mathcal{B}$ such that $T_g A = A$ for all $g \in \mathbb{Z}^d$, we have that $\mu(A) \in \{0, 1\}$.

We say that $(Y, \mathcal{D}, \nu, (S_g)_{g\in\mathbb{Z}^d})$ is a *factor* of $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ if there exists a measurable map $\pi \colon (X, \mathcal{B}, \mu) \to (Y, \mathcal{D}, \nu)$ such that $\mu(\pi^{-1}(A)) = \nu(A)$ for all $A \in \mathcal{D}$, and that $\pi \circ T_g = S_g \circ \pi$ for all $g \in \mathbb{Z}^d$. A factor $(Y, \mathcal{D}, \nu, (S_g)_{g\in\mathbb{Z}^d})$ of $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ can be identified as a sub-σ-algebra $\mathcal{B}'$ of $\mathcal{B}$ or a subspace V of $L^2(\mu)$ by setting $\mathcal{B}' := \pi^{-1}(\mathcal{D})$ or $V := L^2(\nu) \circ \pi$. Given two σ-algebras $\mathcal{B}_1$ and $\mathcal{B}_2$, their *joining* $\mathcal{B}_1 \vee \mathcal{B}_2$ is the σ-algebra generated by $B_1 \cap B_2$ for all $B_1 \in \mathcal{B}_1$ and $B_2 \in \mathcal{B}_2$, *i.e.*, the smallest σ-algebra containing both $\mathcal{B}_1$ and $\mathcal{B}_2$. This definition extends to a countable collection of σ-algebras $\mathcal{B}_i$, $i \in \mathbb{N}$ and we denote it by $\bigvee_{i=0}^{\infty} \mathcal{B}_i$.

We will denote with e_i the vector which has 1 as its ith coordinate and 0's elsewhere. We use in general lower-case letters to symbolize both numbers and vectors but bold letters to symbolize vectors of vectors to highlight this exact fact, in order to make the content more reader-friendly. The only exception to this convention is the vector $\mathbf{0}$ (*i.e.*, the vector with coordinates only 0's) which we always symbolize in bold.

1.4.1. *Notation on averaging.* Throughout this article, we use the following notations about averages. Let $(a(n))_{n\in\mathbb{Z}^L}$ be a sequence of real numbers, or a sequence of measurable functions on a probability space $(X, \mathcal{B}, \mu)$. Denote

$$
\begin{aligned}
\mathbb{E}_{n\in A}\, a(n) &:= \frac{1}{|A|}\sum_{n\in A} a(n), \quad \text{where A is a finite subset of } \mathbb{Z}^L,\\
\overline{\mathbb{E}}^{\square}_{n\in\mathbb{Z}^L}\, a(n) &:= \overline{\lim_{N\to\infty}}\, \mathbb{E}_{n\in[-N,N]^L}\, a(n),^{12}\\
\overline{\mathbb{E}}_{n\in\mathbb{Z}^L}\, a(n) &:= \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}}\, \mathbb{E}_{n\in I_N}\, a(n),\\
\mathbb{E}^{\square}_{n\in\mathbb{Z}^L}\, a(n) &:= \lim_{N\to\infty} \mathbb{E}_{n\in[-N,N]^L}\, a(n) \quad \text{(provided that the limit exists)},\\
\mathbb{E}_{n\in\mathbb{Z}^L}\, a(n) &:= \lim_{N\to\infty} \mathbb{E}_{n\in I_N}\, a(n) \text{ (provided that the limit exists for all Følner sequences } (I_N)_{N\in\mathbb{N}}).
\end{aligned}
$$

We also consider *iterated* averages. Let $(a(h_1, \dots, h_s))_{h_1,\dots,h_s\in\mathbb{Z}^L}$ be a multi-parameter sequence. We denote

$$\overline{\mathbb{E}}_{h_1,\dots,h_s\in\mathbb{Z}^L}\, a(h_1, \dots, h_s) := \overline{\mathbb{E}}_{h_1\in\mathbb{Z}^L} \dots \overline{\mathbb{E}}_{h_s\in\mathbb{Z}^L}\, a(h_1, \dots, h_s)$$

and adopt similar conventions for $\mathbb{E}_{h_1,\dots,h_s\in\mathbb{Z}^L}$, $\overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s\in\mathbb{Z}^L}$ and $\mathbb{E}^{\square}_{h_1,\dots,h_s\in\mathbb{Z}^L}$.

Convention. Throughout this paper, all the limits of measurable functions on a measure preserving system are taken in L^2 (unless otherwise stated). Even though all the expressions with

[12] We use the symbol $\square$ to highlight the fact that the average is along the boxes $[-N, N]^L$.

polynomial iterates that we will encounter converge (in L^2) by [25], we don't a priori postulate any existence of such limits throughout the whole article.

2. Background material

2.1. **The van der Corput lemma.** The main tool in reducing the complexity of polynomial families and running the PET induction is the van der Corput lemma (and its variations), whose original proof can be found in [6]. We state a convenient for us version that can be easily deduced from the one in [6].

Lemma 2.1 ([6]). *Let $\mathcal{H}$ be a Hilbert space, $a\colon \mathbb{Z}^L \to \mathcal{H}$ be a sequence bounded by 1, and $(I_N)_{N\in\mathbb{N}}$ be a Følner sequence in $\mathbb{Z}^L$. Then*

$$\overline{\lim_{N\to\infty}} \|\mathbb{E}_{n\in I_N} a(n)\|^2 \leq 4\overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \overline{\lim_{N\to\infty}} |\mathbb{E}_{n\in I_N} \langle a(n+h), a(n)\rangle|.$$

We also need the following variation of Lemma 2.1:

Lemma 2.2. *Let $\mathcal{H}$ be a Hilbert space, $(a(n;h_1,\dots,h_s))_{(n;h_1,\dots,h_s)\in(\mathbb{Z}^L)^{s+1}}$[13] be a sequence bounded by 1 in $\mathcal{H}$, and $(I_N)_{N\in\mathbb{N}}$ be a Følner sequence in $\mathbb{Z}^L$. Then for $\kappa\in\mathbb{N}$,*

$$\begin{aligned} &\overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \|\mathbb{E}_{n\in I_N} a(n;h_1,\dots,h_s)\|^{2\kappa} \\ &\leq 4^\kappa \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s,h_{s+1}\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} |\mathbb{E}_{n\in I_N} \langle a(n+h_{s+1};h_1,\dots,h_s), a(n;h_1,\dots,h_s)\rangle|^\kappa \end{aligned}$$

Proof. For fixed $h_1,\dots,h_s$, we apply Lemma 2.1 for $a(n)=a(n;h_1,\dots,h_s)$ and $h=h_{s+1}$. By Jensen's inequality, we have

$$\begin{aligned} &\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \|\mathbb{E}_{n\in I_N} a(n;h_1,\dots,h_s)\|^{2\kappa} \\ &\leq 4^\kappa \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \left(\overline{\mathbb{E}}^{\square}_{h_{s+1}\in\mathbb{Z}^L} \overline{\lim_{N\to\infty}} |\mathbb{E}_{n\in I_N} \langle a(n+h_{s+1};h_1,\dots,h_s), a(n;h_1,\dots,h_s)\rangle|\right)^\kappa \\ &\leq 4^\kappa \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\mathbb{E}}^{\square}_{h_{s+1}\in\mathbb{Z}^L} \overline{\lim_{N\to\infty}} |\mathbb{E}_{n\in I_N} \langle a(n+h_{s+1};h_1,\dots,h_s), a(n;h_1,\dots,h_s)\rangle|^\kappa \\ &\leq 4^\kappa \overline{\mathbb{E}}^{\square}_{h_{s+1}\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} |\mathbb{E}_{n\in I_N} \langle a(n+h_{s+1};h_1,\dots,h_s), a(n;h_1,\dots,h_s)\rangle|^\kappa . \end{aligned}$$

The conclusion follows by taking the limsup of the averages over $h_s,\dots,h_1$. □

[13] We use this unorthodox notation to separate the variable n from the h_i's. The variable n plays a different role later.

2.2. **Host-Kra characteristic factors.** The use of Host-Kra characteristic factors is a fundamental tool in studying problems related to multiple averages. They were first introduced in [16] for ergodic $\mathbb{Z}$-systems (see also [27]) and later for $\mathbb{Z}^d$-systems in [15]. In this paper, we need to use a slightly more general version of Host-Kra characteristic factors, which is similar to the one used in [22].

For a $\mathbb{Z}^d$-measure preserving system $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ and a subgroup H of $\mathbb{Z}^d$, $\mathcal{I}(H)$ denotes the sub-σ-algebra of $(T_h)_{h\in H}$-invariant sets, *i.e.*, sets $A \in \mathcal{B}$ such that $T_h A = A$ for all $h \in H$. Let $\mathcal{A}$ be an invariant sub-σ-algebra of $\mathcal{B}$, the measure $\mu \times_{\mathcal{A}} \mu$ denotes the *relative independent product of μ with itself over $\mathcal{A}$*. That is, $\mu \times_{\mathcal{A}} \mu$ is the measure defined on the product space $X \times X$ as

$$\int_{X\times X} f \otimes g \; d(\mu \times_{\mathcal{A}} \mu) = \int_X \mathbb{E}(f|\mathcal{A})\mathbb{E}(g|\mathcal{A}) d\mu$$

for all $f, g \in L^\infty(\mu)$.

Let $H_1, \dots, H_k$ be subgroups of $\mathbb{Z}^d$. Define

$$\mu_{H_1} = \mu \times_{\mathcal{I}(H_1)} \mu$$

and for $k > 1$,

$$\mu_{H_1,\dots,H_k} = \mu_{H_1,\dots,H_{k-1}} \times_{\mathcal{I}(H_k^{[k-1]})} \mu_{H_1,\dots,H_{k-1}},$$

where $H_k^{[k-1]}$ denotes the subgroup of $(\mathbb{Z}^d)^{2^{k-1}}$ consisting of all the elements of the form $h_k \times \dots \times h_k$ (2^{k-1} copies of h_k) for some $h_k \in H_k$. Define the characteristic factor $Z_{H_1,\dots,H_k}(\mathbf{X})$ (or $Z_{H_1,\dots,H_k}$ when there is no confusion) to be the sub-σ-algebra of $\mathcal{B}$ such that

$$\mathbb{E}(f|Z_{H_1,\dots,H_k}) = 0 \text{ if and only if } \|f\|_{H_1,\dots,H_k}^{2^k} := \int_{X^{[k]}} f^{\otimes 2^k} \, d\mu_{H_1,\dots,H_k} = 0,$$

where $f^{\otimes 2^k} = f \otimes \dots \otimes f$ and $X^{[k]} = X \times \dots \times X$ (2^k copies of f and X respectively). Similar to the proof of Lemma 4 of [15] (or Lemma 4.3 of [16]), one can show that $Z_{H_1,\dots,H_k}$ is well defined. Note that when $k = 1$, $Z_{H_1} = \mathcal{I}(H_1)$. When we have k copies of H, we write $Z_{H^{\times k}} := Z_{H,\dots,H}$, and $Z_{H^{\times\infty}} := \bigvee_{k=1}^{\infty} Z_{H^{\times k}}$.

Convention. For convenience, we adopt a flexible way to write the Host-Kra characteristic factors combining the aforementioned notation. For example, if $A = \{H_1, H_2\}$, then the notation $Z_{A,H_3,H_4^{\times 2},(H_i)_{i=5,6}}$ refers to $Z_{H_1,H_2,H_3,H_4,H_4,H_5,H_6}$, and $Z_{H_1,H_2^{\times\infty},H_3^{\times\infty}}$ refers to $\bigvee_{k=1}^{\infty} Z_{H_1,H_2^{\times k},H_3^{\times k}}$.[14] We adopt a similar flexibility for the subscripts of the seminorms.

When each H_i is generated by a single element g_i, we write $\|\cdot\|_{g_1,\dots,g_d} := \|\cdot\|_{H_1,\dots,H_d}$ and $Z_{g_1,\dots,g_d} := Z_{H_1,\dots,H_d}$ in short.

For the rest of the section, $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ will denote, as usual, a $\mathbb{Z}^d$-system.

[14] Or, equivalently $\bigvee_{k_1=1}^{\infty} \bigvee_{k_2=1}^{\infty} Z_{H_1,H_2^{\times k_1},H_3^{\times k_2}}$.

Let H be a subgroup of $\mathbb{Z}^d$ and $(a(g))_{g\in H}$ be a sequence on a Hilbert space. If for all Følner sequences $(I_N)_{N\in\mathbb{N}}$ of H, the limit $\lim_{N\to\infty}\mathbb{E}_{g\in I_N}a(g)$ exists, we then use $\mathbb{E}_{g\in H}a(g)$ to denote this limit.[15] The following theorem is classical (see for example [10, Theorem 8.13]):

Theorem 2.3 (Mean ergodic theorem for $\mathbb{Z}^d$-actions)**.** *For every $f\in L^2(\mu)$ and every subgroup H of $\mathbb{Z}^d$, the limit $\mathbb{E}_{g\in H}T_gf$ exists in $L^2(\mu)$ and equals to $\mathbb{E}(f|\mathcal{I}(H))$ (or $\mathbb{E}(f|Z_H)$).*

The following are some basic properties of the Host-Kra seminorms:

Lemma 2.4. *Let $H_1,\dots,H_k,H'$ be subgroups of $\mathbb{Z}^d$ and $f\in L^\infty(\mu)$.*

(i) *For every permutation $\sigma\colon\{1,\dots,k\}\to\{1,\dots,k\}$, we have that*
$$Z_{H_1,\dots,H_k}(\mathbf{X})=Z_{H_{\sigma(1)},\dots,H_{\sigma(k)}}(\mathbf{X}).$$
(ii) *If $\mathcal{I}(H_j)=\mathcal{I}(H')$, then $Z_{H_1,\dots,H_j,\dots,H_k}(\mathbf{X})=Z_{H_1,\dots,H_{j-1},H',H_{j+1},\dots,H_k}(\mathbf{X})$*

(iii) *For $k\geq 2$ we have that*
$$\|f\|^{2^k}_{H_1,\dots,H_k}=\mathbb{E}_{g\in H_k}\Big\|f\cdot T_gf\Big\|^{2^{k-1}}_{H_1,\dots,H_{k-1}},$$
while for $k=1$,
$$\|f\|^2_{H_1}=\mathbb{E}_{g\in H_1}\int_X f\cdot T_gf\,d\mu.$$
(iv) *Let $k\geq 2$. If $H'\leq H_j$ is of finite index, then*
$$Z_{H_1,\dots,H_j,\dots,H_k}(\mathbf{X})=Z_{H_1,\dots,H_{j-1},H',H_{j+1},\dots,H_k}(\mathbf{X}).$$
(v) *If $H'\leq H_j$, then $Z_{H_1,\dots,H_j,\dots,H_k}(\mathbf{X})\subseteq Z_{H_1,\dots,H_{j-1},H',H_{j+1},\dots,H_k}(\mathbf{X})$.*

(vi) *For $k\geq 1$, $\|f\|_{H_1,\dots,H_{k-1}}\leq\|f\|_{H_1,\dots,H_{k-1},H_k}$ and thus $Z_{H_1,\dots,H_{k-1}}(\mathbf{X})\subseteq Z_{H_1,\dots,H_{k-1},H_k}(\mathbf{X})$.*

(vii) *For $k\geq 1$, if $H_1',\dots,H_k'$ are subgroups of $\mathbb{Z}^d$, then $Z_{H_1,\dots,H_k}(\mathbf{X})\vee Z_{H_1',\dots,H_k'}(\mathbf{X})\subseteq Z_{H_1',\dots,H_k',H_1,\dots,H_k}(\mathbf{X})$.*

Proof. (i) and (ii) follow from [22, Lemma 2.2] (for (i), see also [15]).

To show (iii), if $k\geq 2$, then
$$\begin{aligned}\|f\|^{2^k}_{H_1,\dots,H_k} &= \int_{X^{[k]}} f^{\otimes 2^k}\,d\mu_{H_1,\dots,H_k}\\ &= \int_{X^{[k-1]}} f^{\otimes 2^{k-1}}\cdot\mathbb{E}(f^{\otimes 2^{k-1}}|\mathcal{I}(H_k^{[k-1]}))\,d\mu_{H_1,\dots,H_{k-1}}\\ &= \mathbb{E}_{g\in H_k}\int_{X^{[d-1]}} f^{\otimes 2^{k-1}}\cdot(T_gf)^{\otimes 2^{k-1}}\,d\mu_{H_1,\dots,H_{k-1}}\\ &= \mathbb{E}_{g\in H_k}\Big\|f\cdot T_gf\Big\|^{2^{k-1}}_{\mathbf{X},H_1,\dots,H_{k-1}},\end{aligned}$$
where we invoke the mean ergodic theorem (Theorem 2.3) in the third equality. Similarly, for $k=1$,
$$\|f\|^2_{H_1}=\int_{X^2} f\otimes f\,d\mu_{H_1}=\int_X f\cdot\mathbb{E}(f|\mathcal{I}(H_1))\,d\mu=\mathbb{E}_{g\in H_1}\int_X f\cdot T_gf\,d\mu.$$

[15] The fact that the limit exists for all Følner sequences actually implies that the limit is the same for all of them.

We now prove (iv). By (i), we may assume without loss of generality that $j = k$. Suppose that $H_k = \sqcup_{i=1}^{l} g_i H'$ for some $l > 0$ and $g_i \in \mathbb{Z}^d, 1 \leq i \leq l$. We may assume that $g_1 = e_G$. Let $(I_N)_{N\in\mathbb{N}}$ be any Følner sequence on H'. We claim that $(I_N \cdot \{g_1, \dots, g_l\})_{N\in\mathbb{N}}$ is a Følner sequence in H_k. Indeed, by the elementary inclusion $(A \cup B)\triangle C \subseteq (A\triangle C) \cup (B\triangle C)$ it follows that [16]

$$(I_N \cdot \{g_1, \dots, g_l\})\triangle g(I_N \cdot \{g_1, \dots, g_l\}) \subseteq \bigcup_{1\leq i,j\leq l} I_N g_i \triangle g I_N g_j = \bigcup_{1\leq i,j\leq l} g_i I_N \triangle g_j g I_N,$$

and since $|I_N|^{-1} \cdot |g_i I_N \triangle g_j g I_N| = |I_N|^{-1} \cdot |I_N \triangle (g_i^{-1} g_j g) I_N| \to 0$ as $N \to \infty$, the claim follows.

By (iii), we have that

$$(5)\quad \begin{aligned} \|f\|_{H_1,\dots,H_{k-1},H_k}^{2^k} &= \mathbb{E}_{g\in H_k} \Big\| f \cdot T_g f \Big\|_{H_1,\dots,H_{k-1}}^{2^{k-1}} \\ &= \lim_{N\to\infty} \frac{1}{l|I_N|} \sum_{i=1}^{l} \sum_{g\in I_N} \Big\| f \cdot T_{g_i g} f \Big\|_{H_1,\dots,H_{k-1}}^{2^{k-1}} \\ &\geq \lim_{N\to\infty} \frac{1}{l|I_N|} \sum_{g\in I_N} \Big\| f \cdot T_g f \Big\|_{H_1,\dots,H_{k-1}}^{2^{k-1}} \\ &= \frac{1}{l} \|f\|_{H_1,\dots,H_{k-1},H'}^{2^k}. \end{aligned}$$

On the other hand, since $\mathcal{I}(H_k^{[k-1]})$ is a sub σ-algebra of $\mathcal{I}(H'^{[k-1]})$, by the Cauchy-Schwarz inequality,

$$(6)\quad \begin{aligned} \|f\|_{H_1,\dots,H_{k-1},H'}^{2^k} &= \int_{X^{[k]}} f^{\otimes 2^k} \, d\mu_{H_1,\dots,H_{k-1},H'} \\ &= \int_{X^{[k-1]}} f^{\otimes 2^{k-1}} \cdot \mathbb{E}(f^{\otimes 2^{k-1}} | \mathcal{I}(H'^{[k-1]})) \, d\mu_{H_1,\dots,H_{k-1}} \\ &= \int_{X^{[k-1]}} \Big| \mathbb{E}(f^{\otimes 2^{k-1}} | \mathcal{I}(H'^{[k-1]})) \Big|^2 \, d\mu_{H_1,\dots,H_{k-1}} \\ &\geq \int_{X^{[k-1]}} \Big| \mathbb{E}(f^{\otimes 2^{k-1}} | \mathcal{I}(H_k^{[k-1]})) \Big|^2 \, d\mu_{H_1,\dots,H_{k-1}} \\ &= \int_{X^{[k-1]}} f^{\otimes 2^{k-1}} \cdot \mathbb{E}(f^{\otimes 2^{k-1}} | \mathcal{I}(H_k^{[k-1]})) \, d\mu_{H_1,\dots,H_{k-1}} \\ &= \int_{X^{[k]}} f^{\otimes 2^k} \, d\mu_{H_1,\dots,H_{k-1},H_k} = \|f\|_{H_1,\dots,H_{k-1},H_k}^{2^k}. \end{aligned}$$

Therefore, $\|f\|_{H_1,\dots,H_{k-1},H_k} = 0 \Leftrightarrow \|f\|_{H_1,\dots,H_{k-1},H'} = 0$, and the conclusion follows.

(v) Since $\|f\|_{H_1,\dots,H_{k-1},H_k}^{2^k} \leq \|f\|_{H_1,\dots,H_{k-1},H'}^{2^k}$ by (6) whenever H' is a subgroup of H_k, we have that $Z_{H_1,\dots,H_{k-1},H_k}(\mathbf{X}) \subseteq Z_{H_1,\dots,H_{k-1},H'}(\mathbf{X})$. So (v) follows from (i).

[16] We use multiplicative notation for convenience.

(vi) Similarly to (iii), and by Jensen inequality we have

$$\begin{aligned}
\|f\|_{H_1,\dots,H_{k-1},H_k}^{2^k} &= \int_{X^{[k]}} f^{\otimes 2^k}\, d\mu_{H_1,\dots,H_k} \\
&= \int_{X^{[k-1]}} \mathbb{E}(f^{\otimes 2^{k-1}}|\mathcal{I}(H_k^{[k-1]}))^2\, d\mu_{H_1,\dots,H_{k-1}} \\
&\geq \Big(\int_{X^{[k-1]}} \mathbb{E}(f^{\otimes 2^{k-1}}|\mathcal{I}(H_k^{[k-1]}))\, d\mu_{H_1,\dots,H_{k-1}}\Big)^2 \\
&= \Big(\int_{X^{[k-1]}} f^{\otimes 2^{k-1}}\, d\mu_{H_1,\dots,H_{k-1}}\Big)^2 \\
&= \|f\|_{H_1,\dots,H_{k-1}}^{2^k},
\end{aligned}$$

from where the conclusion follows.

(vii) Applying (vi) several times, we get that both $Z_{H_1,\dots,H_k}(\mathbf{X})$ and$Z_{H'_1,\dots,H'_k}(\mathbf{X})$ are sub-σ-algebras of $Z_{H'_1,\dots,H'_k,H_1,\dots,H_k}(\mathbf{X})$, hence so is their joining.

□

Remark. We caution the reader that Lemma 2.4 (iii) is not valid for $k = 1$. In fact, for an ergodic $\mathbb{Z}$-system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ where T^2 is not ergodic, we have $Z_{\mathbb{Z}}(\mathbf{X}) = I(\mathbb{Z}) \neq I(2\mathbb{Z}) = Z_{2\mathbb{Z}}(\mathbf{X})$. The reason that (iii) fails for $k = 1$ is that the inequality in (5) is no longer valid since the term $\Big\|f \cdot T_{g_i g} f\Big\|_{\mathbf{X},H_1,\dots,H_{k-1}}^{2^{k-1}}$ is replaced by $\int_X f \cdot T_{g_i g} f\, d\mu$, which might be negative.

As an immediate corollary of Lemma 2.4 (ii), we have:

Corollary 2.5. *Let $H_1, \dots, H_k$ be subgroups of $\mathbb{Z}^d$. If the H_i-action $(T_g)_{g\in H_i}$ is ergodic on $\mathbf{X}$ for all $1 \leq i \leq k$, then $Z_{H_1,\dots,H_k}(\mathbf{X}) = Z_{(\mathbb{Z}^d)^{\times k}}(\mathbf{X})$.*

2.3. **Structure theorem and nilsystems.** Let $X = N/\Gamma$, where N is a (k-step) nilpotent Lie group and Γ is a discrete cocompact subgroup of N. Let $\mathcal{B}$ be the Borel σ-algebra of X, μ the Haar measure on X, and for $g \in \mathbb{Z}^d$, let $T_g\colon X \to X$ with $T_g x = b_g \cdot x$ for some group homomorphism $g \mapsto b_g$ from $\mathbb{Z}^d$ to N. We say that $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ is a *(k-step) $\mathbb{Z}^d$-nilsystem.*

An important reason which makes the Host-Kra characteristic factors powerful is their connection with nilsystems. The following is a slight generalization of [26, Theorem 3.7] (see [22, Theorem 3.7]), which is a higher dimensional version of Host-Kra structure theorem ([16]).

Theorem 2.6 (Structure theorem)**.** *Let $\mathbf{X}$ be an ergodic $\mathbb{Z}^d$-system. Then $Z_{(\mathbb{Z}^d)^{\times k}}(\mathbf{X})$ is an inverse limit of $(k-1)$-step $\mathbb{Z}^d$-nilsystems.*

The 1-step Host-Kra nilfactor is the *Kronecker factor*, which is intimately related to the spectrum of the system [16]. We say that a non-μ-a.e. constant function $f \in L^\infty(\mu)$ is an *eigenfunction* of the $\mathbb{Z}^d$-system $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ if $T_g f = \lambda_g f$ for all $g \in \mathbb{Z}^d$, where $g \mapsto \lambda_g$ is a group homomorphism from $\mathbb{Z}^d$ to $\mathbb{S}^1$. For each $g \in \mathbb{Z}^d$, we say that λ_g is an *eigenvalue* of $\mathbf{X}$. If $(X, \mathcal{B}, \mu, T)$ is a $\mathbb{Z}$-system, we say that a non-μ-a.e. constant function $f \in L^\infty(\mu)$ is an *eigenfunction* of T if $Tf = \lambda f$ for some $\lambda \in \mathbb{S}^1$, and we say that λ is an *eigenvalue* of T.

The *Kronecker factor* $\mathcal{K}(\mathbf{X})$ of the $\mathbb{Z}^d$-system $\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ is the sub-σ-algebra of $\mathcal{B}$ that corresponds to the algebra of functions spanned by the eigenfunctions of $\mathbf{X}$ in $L^2(\mu)$. As a special case of Theorem 2.6, we have:

Lemma 2.7. *For an ergodic $\mathbb{Z}^d$-system $\mathbf{X}$, we have that $\mathcal{K}(\mathbf{X}) = Z_{\mathbb{Z}^d,\mathbb{Z}^d}(\mathbf{X})$.*

An application of the Kronecker factor is to characterize single averages along polynomials (for a proof of this result, see Section 2 in [5]):

Proposition 2.8. *Let $L \in \mathbb{N}^*$, $p\colon \mathbb{Z}^L \to \mathbb{Z}$ be a non-constant polynomial,*[17] *$\mathbf{X} = (X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}})$ be a $\mathbb{Z}$-system, and $f \in L^\infty(\mu)$. If $\mathbb{E}(f|Z_{\mathbb{Z},\mathbb{Z}}(\mathbf{X})) = 0$, then*

$$\mathbb{E}_{n\in\mathbb{Z}^L} T_{p(n)} f = 0.$$

We provide an alternative proof of Proposition 2.8 in Section 4 using the language of this paper.

We conclude this subsection with the following theorem from [20], a consequence of [20, Theorem B], which we state in a convenient form.

Theorem 2.9 (Leibman, [20])**.** *Let $(X = N/\Gamma, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ be a nilsystem, $x \in X$ and $p_i\colon \mathbb{Z}^L \to \mathbb{Z}^d$, $1 \le i \le k$ be polynomials. Then the following are equivalent:*

(i) *For some $x \in X$, the sequence $(T_{p_1(n)}x, \dots, T_{p_k(n)}x)_{n\in\mathbb{Z}^L}$ is dense in X^k.*
(ii) *The sequence $(T_{p_1(n)}y_1, \dots, T_{p_k(n)}y_k)_{n\in\mathbb{Z}^L}$ is dense in X^k for all $y_1, \dots, y_k \in X$.*
(iii) *For any Følner sequence $(I_N)_{N\in\mathbb{N}}$ in $\mathbb{Z}^L$, the average $\mathbb{E}_{n\in I_N} f_1(T_{p_1(n)}x) \cdot \dots \cdot f_k(T_{p_k(n)}x)$ converges (in L^2 and μ-a.e.) to $\prod_{i=1}^{k} \int f_i d\mu$ as $N \to \infty$.*
(iv) *The sequence $(T_{p_1(n)}, \dots, T_{p_k(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes k}$.*

2.4. **Concatenation theorem.** An essential ingredient in our approach is the concatenation theorem established by Tao and Ziegler (in [24]), which studies the properties of intersections of different characteristic factors:

Theorem 2.10 (Concatenation theorem, [24, Theorem 1.15])**.** *Let $\mathbf{X}$ be a $\mathbb{Z}^d$-system, $k, k' \in \mathbb{N}^*$ and $H_1, \dots, H_k, H'_1, \dots, H'_{k'}$ subgroups of $\mathbb{Z}^d$. Then*

$$Z_{H_1,\dots,H_k} \cap Z_{H'_1,\dots,H'_{k'}} \subseteq Z_{(H_i+H'_{i'})_{1\le i\le k, 1\le i'\le k'}}.$$

As an immediate corollary, we have:

Corollary 2.11. *Let $\mathbf{X}$ be a $\mathbb{Z}^d$-system, $s, d_1, \dots, d_s \in \mathbb{N}$ and $H_{i,j}, 1 \le i \le s, 1 \le j \le d_i$ subgroups of $\mathbb{Z}^d$. Then*

$$\bigcap_{i=1}^{s} Z_{H_{i,1},H_{i,2},\dots,H_{i,d_i}} \subseteq Z_{(H_{1,n_1}+H_{2,n_2}+\dots+H_{s,n_s})_{1\le n_i\le d_i, 1\le i\le s}}.$$

2.5. **Range of polynomials.** In this subsection we state and prove two elementary lemmas regarding the range of polynomials.

Definition. For $\mathbf{b} = (b_1, \dots, b_L) \in (\mathbb{Q}^d)^L, b_i \in \mathbb{Q}^d$, we define

$$G(\mathbf{b}) := \mathrm{span}_{\mathbb{Q}}\{b_1, \dots, b_L\} \cap \mathbb{Z}^d. \tag{7}$$

Note that $G(\mathbf{b})$ can either be seen as a subgroup or a subspace (over $\mathbb{Z}$) of $\mathbb{Z}^d$; we freely use both.

[17] We warn the reader that this result is only valid for $p\colon \mathbb{Z}^L \to \mathbb{Z}^d$ with d equal to 1.

Lemma 2.12. *Let* $\mathbf{c}\colon (\mathbb{Z}^L)^s \to (\mathbb{Q}^d)^L$ *be a polynomial and let* V *be a subspace of* $\mathbb{Z}^d$ *over* $\mathbb{Z}$. *Then the set*

$$\{(h_1,\dots,h_s)\in(\mathbb{Z}^L)^s\colon G(\mathbf{c}(h_1,\dots,h_s))\subseteq V\}$$

is either $(\mathbb{Z}^L)^s$ *or of (upper) Banach density* 0.

Proof. For convenience, denote

$$W := \{(h_1,\dots,h_s)\in(\mathbb{Z}^L)^s\colon G(\mathbf{c}(h_1,\dots,h_s))\subseteq V\},$$

where one views $\mathbf{c}$ as the matrix:

$$\mathbf{c}(h_1,\dots,h_s)=\begin{pmatrix} c_{1,1}(h_1,\dots,h_s) & \dots & c_{1,L}(h_1,\dots,h_s)\\ \vdots & \vdots & \vdots \\ c_{d,1}(h_1,\dots,h_s) & \dots & c_{d,L}(h_1,\dots,h_s)\end{pmatrix}$$

for some polynomials $c_{i,j}\colon(\mathbb{Z}^L)^s\to\mathbb{Q}$, $1\le i\le d, 1\le j\le L$.

We start with the case $V=\{\mathbf{0}\}$. Let $W_{i,j}$ be the set of $(h_1,\dots,h_s)\in(\mathbb{Z}^L)^s$ such that $c_{i,j}(h_1,\dots,h_s)=0$. Then $W=\bigcap_{i=1}^{d}\bigcap_{j=1}^{L}W_{i,j}$ and so it suffices to show that either each $W_{i,j}$ is $(\mathbb{Z}^L)^s$ or that some $W_{i,j}$ is of density 0. By relabelling the variables, we may assume that $L=1$ (and change s to Ls). Hence, it suffices to show that for a polynomial $c\colon\mathbb{Z}^s\to\mathbb{Z}$, the set

$$W=\{(h_1,\dots,h_s)\in(\mathbb{Z})^s\colon c(h_1,\dots,h_s)=0\}$$

is either $\mathbb{Z}^s$ or of density 0.

If $s=1$, then either $c\equiv 0$ or $c(x)=0$ has finitely many roots. So W is either $\mathbb{Z}$ or of upper Banach density 0. Suppose now that the conclusion holds for some $s\ge 1$, and assume that $c(h_1,\dots,h_{s+1})=\sum_{i=0}^{K}q_i(h_2,\dots,h_{s+1})h_1^i$ for some $K\in\mathbb{N}$ and polynomials $q_i\colon\mathbb{Z}^s\to\mathbb{Q}$ for all $0\le i\le K$. Let

$$W'=\{(h_2,\dots,h_{s+1})\in\mathbb{Z}^s\colon q_i(h_2,\dots,h_{s+1})=0, 0\le i\le K\}.$$

By induction hypothesis, either $W'=\mathbb{Z}^s$ or W' is of upper Banach density 0. If $W'=\mathbb{Z}^s$, then $c\equiv 0$ and so $W=\mathbb{Z}^{s+1}$. If W' is of upper Banach density 0, then $W\subseteq W_1\cup W_2$, where $W_1=\mathbb{Z}\times W'$ and $W_2=\{(h_1,\dots,h_{s+1})\in\mathbb{Z}^{s+1}\colon (h_2,\dots,h_{s+1})\notin W', c(h_1,\dots,h_{s+1})=0\}$. Since W' is of upper Banach density 0, so is W_1. On the other hand, for any $(h_2,\dots,h_{s+1})\notin W'$, $c(\cdot,h_2,\dots,h_{s+1})$ is not constant 0 and so has at most K roots. This implies that W_2 is of upper Banach density 0, so W is of density 0, completing the induction.

Now assume that $V\neq\{\mathbf{0}\}$. Since V is a subspace of $\mathbb{Z}^d$ over $\mathbb{Z}$, under a change of coordinates, we may assume that $V=\{0\}^\ell\times\mathbb{Z}^{d-\ell}$ for some $0\le\ell\le d$. If $\ell=0$, then $V=\mathbb{Z}^d$ and there is nothing to prove. If $\ell>0$, then by restricting to the first polynomials $c_{i,j}, 1\le i\le d, 1\le j\le\ell$, we are reduced to the case $V=\{\mathbf{0}\}$, finishing the proof. □

Lemma 2.13. *Let* $\mathbf{c}\colon(\mathbb{Z}^L)^s\to(\mathbb{Q}^d)^L$ *be a polynomial given by*

$$\mathbf{c}(h_1,\dots,h_s)=\sum_{a_1,\dots,a_s\in\mathbb{N}^L}h_1^{a_1}\dots h_s^{a_s}\cdot\mathbf{u}(a_1,\dots,a_s)^{18}$$

[18] Recall that for $n=(n_1,\dots,n_L)\in\mathbb{Z}^L$ and $v=(v_1,\dots,v_L)\in\mathbb{N}^L$, n^v denotes the quantity $n_1^{v_1}\dots n_L^{v_L}$. We also use the convention $0^0=1$.

for some $\mathbf{u}(a_1,\ldots,a_s)\in(\mathbb{Q}^d)^L$ *which all but finitely many equal to* 0. *Then*

$$\mathit{span}_{\mathbb{Q}}\{G(\mathbf{c}(h_1,\ldots,h_s))\colon h_1,\ldots,h_s\in\mathbb{Z}^L\}=\mathit{span}_{\mathbb{Q}}\{G(\mathbf{u}(a_1,\ldots,a_s))\colon a_1,\ldots,a_s\in\mathbb{N}^L\}.$$ [19]

For the reader's convenience we first make the statement clear with an example, with $L=2$, $s=1$, $d=4$, and then present the proof. Let $\mathbf{c}\colon\mathbb{Z}^2\to(\mathbb{Z}^4)^2$ be given by

$$\mathbf{c}(h_1,h_2)=\begin{pmatrix} h_1 & 0\\ -3h_1h_2 & h_1\\ h_1^2 & -h_2-2h_2^2\\ 7h_1h_2 & h_1^2\end{pmatrix}.$$

Denoting $h=(h_1,h_2)$, we have

$$\begin{aligned}\mathbf{c}(h_1,h_2)&=h_1\begin{pmatrix}1&0\\0&1\\0&0\\0&0\end{pmatrix}+h_2\begin{pmatrix}0&0\\0&0\\0&-1\\0&0\end{pmatrix}+h_1h_2\begin{pmatrix}0&0\\-3&0\\0&0\\7&0\end{pmatrix}+h_1^2\begin{pmatrix}0&0\\0&0\\1&0\\0&1\end{pmatrix}+h_2^2\begin{pmatrix}0&0\\0&0\\0&-2\\0&0\end{pmatrix}\\
&=h^{(1,0)}\begin{pmatrix}1&0\\0&1\\0&0\\0&0\end{pmatrix}+h^{(0,1)}\begin{pmatrix}0&0\\0&0\\0&-1\\0&0\end{pmatrix}+h^{(1,1)}\begin{pmatrix}0&0\\-3&0\\0&0\\7&0\end{pmatrix}+h^{(2,0)}\begin{pmatrix}0&0\\0&0\\1&0\\0&1\end{pmatrix}+h^{(0,2)}\begin{pmatrix}0&0\\0&0\\0&-2\\0&0\end{pmatrix}\\
&=h^{(1,0)}\mathbf{u}(1,0)+h^{(0,1)}\mathbf{u}(0,1)+h^{(1,1)}\mathbf{u}(1,1)+h^{(2,0)}\mathbf{u}(2,0)+h^{(0,2)}\mathbf{u}(0,2),\end{aligned}$$

where the $\mathbf{u}(i,j)$ denote the corresponding matrices from the previous step.

Lemma 2.13 establishes that the span of the columns of $\mathbf{c}(h_1,h_2)$ (for all $h_1,h_2\in\mathbb{Z}$) equals to the span of the columns of the $\mathbf{u}(a_1,a_2)$ (for all $a_1,a_2\in\mathbb{N}$). More explicitly, it states that

$$\mathrm{span}_{\mathbb{Q}}\left\{\begin{pmatrix}h_1\\-3h_1h_2\\h_1^2\\7h_1h_2\end{pmatrix},\begin{pmatrix}0\\h_1\\-h_2-2h_2^2\\h_1^2\end{pmatrix}:h_1,h_2\in\mathbb{Z}\right\}$$

equals to

$$\mathrm{span}_{\mathbb{Q}}\left\{\begin{pmatrix}1\\0\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\\0\end{pmatrix},\begin{pmatrix}0\\0\\-1\\0\end{pmatrix},\begin{pmatrix}0\\-3\\0\\7\end{pmatrix},\begin{pmatrix}0\\0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\0\\1\end{pmatrix},\begin{pmatrix}0\\0\\-2\\0\end{pmatrix}\right\}.$$

Proof of Lemma 2.13. We first assume that $L=1$. In this case, we have that

$$c(h_1,\ldots,h_s)=\sum_{a_1,\ldots,a_s\in\mathbb{N}}h_1^{a_1}\ldots h_s^{a_s}\cdot u(a_1,\ldots,a_s)$$

for $h_1,\ldots,h_s\in\mathbb{Z}$ and some $u(a_1,\ldots,a_s)\in\mathbb{Q}^d$. It suffices to show that

$$\mathrm{span}_{\mathbb{Q}}\{c(h_1,\ldots,h_s)\colon h_1,\ldots,h_s\in\mathbb{Z}\}=\mathrm{span}_{\mathbb{Q}}\{u(a_1,\ldots,a_s)\colon a_1,\ldots,a_s\in\mathbb{N}\}.$$

[19] Here, when $H_i, i\in\mathbb{N}$ are subsets of $\mathbb{Q}^d$, we use the notation $\mathrm{span}_{\mathbb{Q}}\{H_i\colon i\in\mathbb{N}\}$ to denote the set $\mathrm{span}_{\mathbb{Q}}\{x\in\mathbb{Q}^d\colon x\in\cup_{i\in\mathbb{N}}H_i\}$.

Since $c(h_1,\dots,h_s)$ belongs to the $\mathbb{Q}$-span of $\{u(a_1,\dots,a_s)\}_{a_1,\dots,a_s\in\mathbb{N}}$, the inclusion "$\subseteq$" is straightforward. We then show the "$\supseteq$" direction. When $s=1$, we have that $c(h_1)=\sum_{i=0}^{K} h_1^i u(i)$ for some $K\in\mathbb{N}$. Since the matrix $(j^i)_{0\leq i,j\leq K}$,[20] is (the transpose of) a Vandermonde matrix, its determinant is non-zero, so each $u(i)$ is a linear combination of $c(0),\dots,c(K)$. Therefore, the conclusion holds for $s=1$.

We now suppose that the conclusion holds for some $s\geq 1$ and we prove it for $s+1$. Write

$$c(h_1,\dots,h_{s+1})=\sum_{a_1,\dots,a_{s+1}\in\mathbb{N}} h_1^{a_1}\dots h_{s+1}^{a_{s+1}}\cdot u(a_1,\dots,a_{s+1})=\sum_{i\in\mathbb{N}} h_{s+1}^i v_i(h_1,\dots,h_s)$$

for some polynomials $v_i\colon\mathbb{Z}^s\to\mathbb{Q}^d$ given by

$$v_i(h_1,\dots,h_s)=\sum_{a_1,\dots,a_s\in\mathbb{N}} h_1^{a_1}\dots h_s^{a_s}\cdot u(a_1,\dots,a_s,i).$$

Since the conclusion holds for $s=1$, we have that for all $h_1,\dots,h_s\in\mathbb{Z}$ and $i\in\mathbb{N}$, $v_i(h_1,\dots,h_s)\in \operatorname{span}_{\mathbb{Q}}\{c(h_1,\dots,h_s,h_{s+1}):\ h_{s+1}\in\mathbb{Z}\}$. Applying the induction hypothesis for s, we have that

$$u(a_1,\dots,a_s,i)\in\operatorname{span}_{\mathbb{Q}}\{v_i(h_1,\dots,h_s)\colon h_1,\dots,h_s\in\mathbb{Z}\}$$

for all $a_1,\dots,a_s,i\in\mathbb{N}$, hence the conclusion holds for $s+1$. By induction, the $L=1$ case is complete.

For the general case, suppose that $\mathbf{c}(h_1,\dots,h_s)=(c_1(h_1,\dots,h_s),\dots,c_L(h_1,\dots,h_s))$ and $\mathbf{u}(a_1,\dots,a_s)=(u_1(a_1,\dots,a_s),\dots,u_L(a_1,\dots,a_s))$, where $c_i\colon(\mathbb{Z}^L)^s\to\mathbb{Q}^d$, $u_i\colon(\mathbb{N}^L)^s\to\mathbb{Q}^d$, $1\leq i\leq L$. Then

$$c_i(h_1,\dots,h_s)=\sum_{a_1,\dots,a_s\in\mathbb{N}^L} h_1^{a_1}\dots h_s^{a_s}\cdot u_i(a_1,\dots,a_s) \tag{8}$$

for all $1\leq i\leq L$. By definition, one easily checks that

$$\operatorname{span}_{\mathbb{Q}}\{G(\mathbf{c}(h_1,\dots,h_s))\colon h_1,\dots,h_s\in\mathbb{Z}^L\}=\operatorname{span}_{\mathbb{Q}}\{c_i(h_1,\dots,h_s)\colon h_1,\dots,h_s\in\mathbb{Z}^L,1\leq i\leq L\},$$

and

$$\operatorname{span}_{\mathbb{Q}}\{G(\mathbf{u}(a_1,\dots,a_s))\colon a_1,\dots,a_s\in\mathbb{N}^L\}=\operatorname{span}_{\mathbb{Q}}\{u_i(a_1,\dots,a_s)\colon a_1,\dots,a_s\in\mathbb{N}^L,1\leq i\leq L\}.$$

So, it suffices to show that for every $1\leq i\leq L$,

$$\operatorname{span}_{\mathbb{Q}}\{c_i(h_1,\dots,h_s)\colon h_1,\dots,h_s\in\mathbb{Z}^L\}=\operatorname{span}_{\mathbb{Q}}\{u_i(a_1,\dots,a_s)\colon a_1,\dots,a_s\in\mathbb{N}^L\},\ \text{ or}$$

$$\operatorname{span}_{\mathbb{Q}}\{c_i(h)\colon h\in\mathbb{Z}^{Ls}\}=\operatorname{span}_{\mathbb{Q}}\{u_i(a)\colon a\in\mathbb{N}^{Ls}\}, \tag{9}$$

by viewing $(h_1,\dots,h_s)$ and $(a_1,\dots,a_s)$ as the Ls-dimensional vectors h and a. Rewriting (8) as

$$c_i(h)=\sum_{a\in\mathbb{N}^{Ls}} h^a\cdot u_i(a),$$

we can apply the conclusion of the case $L'=1$, $s'=Ls$, $d'=d$ and $c_i\colon(\mathbb{Z}^{L'})^{s'}=(\mathbb{Z}^L)^s\to(\mathbb{Z}^{d'})^{L'}=\mathbb{Z}^d$ to show (9). This finishes the proof. □

[20] Recall that we set $0^0:=1$.

3. Equivalent conditions for $((T_1 \times \cdots \times T_d)^{p(n)})_{n\in\mathbb{Z}^L}$ being ergodic

In this short section, we provide equivalent conditions of Property (ii) in Theorem 1.4, *i.e.*, we characterize when $((T_1 \times \cdots \times T_d)^{p(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes d}$.

The following lemma is an implication of [12, Lemma 4.18].

Lemma 3.1. *Let $\mathbf{X}_i = (X_i, \mathcal{B}_i, \mu_i, T_i)$, $1 \leq i \leq d$ be $\mathbb{Z}$-systems. The set of eigenvalues of $T_1 \times \cdots \times T_d$ consists of all numbers of the form $\prod_{i=1}^{d} \lambda_i$, where λ_i is either 1 or an eigenvalue of T_i, where at least one λ_i is an eigenvalue.*

Proof. Suppose first that λ_i is either 1 or an eigenvalue of T_i and that at least one λ_i is an eigenvalue. Then, for all $1 \leq i \leq d$, $T_i f_i = \lambda_i f_i$ for some $f_i \in L^\infty(\mu_i)$, where not all f_i's are μ_i-a.e. constant. Then $(T_1 \times \cdots \times T_d)(f_1 \otimes \cdots \otimes f_d) = \left(\prod_{i=1}^{d} \lambda_i\right)(f_1 \otimes \cdots \otimes f_d)$. Since $f_1 \otimes \cdots \otimes f_d$ is not $(\mu_1 \times \cdots \times \mu_d)$-a.e. constant, $\prod_{i=1}^{d} \lambda_i$ is an eigenvalue of $T_1 \times \cdots \times T_d$.

Conversely, let λ be an eigenvalue of $T_1 \times \cdots \times T_d$ with a corresponding eigenfunction f. By [12, Lemma 4.18], $f = \sum_n c_n f_{1,n} \otimes \cdots \otimes f_{d,n}$, where $c_n \in \mathbb{C}$, $T_i f_{i,n} = \lambda_{i,n} f_{i,n}$ for some $\lambda_{i,n} \in \mathbb{S}^1$ with $\prod_{i=1}^{d} \lambda_{i,n} = \lambda$. Each $\lambda_{i,n}$ is either 1 or an eigenvalue of T_i. Since f is not $(\mu_1 \times \cdots \times \mu_d)$-a.e. constant, some $f_{1,n} \otimes \cdots \otimes f_{d,n}$ is also not $(\mu_1 \times \cdots \times \mu_d)$-a.e. constant. For such n, at least one of $\lambda_{1,n}, \ldots, \lambda_{d,n}$ is an eigenvalue of T_i. Note that if $f_{i,n}$ is μ_i-a.e. constant, then $\lambda_{i,n} = 1$. Otherwise $\lambda_{i,n}$ is an eigenvalue for T_i, which finishes the proof. □

Let $p\colon \mathbb{Z}^L \to \mathbb{Z}$ be a polynomial and $\lambda \in \mathbb{S}^1$. We say that λ is *uniform* for p if $\mathbb{E}_{n\in\mathbb{Z}^L} \lambda^{p(n)} = 0$. So, $\lambda = 1$ is not uniform for any integer-valued polynomial, while by Weyl's equidistribution theorem, every $\lambda = e^{2\pi i a}$ for some $a \notin \mathbb{Q}$ is uniform for all integer-valued polynomials.

The following proposition, which lists conditions equivalent to Property (ii) of Theorem 1.4, is the main result of the section.

Proposition 3.2 (Conditions equivalent to (ii) of Theorem 1.4)**.** *Let $(X, \mathcal{B}, \mu, T_1, \ldots, T_d)$ be a system with commuting transformations and $p\colon \mathbb{Z}^L \to \mathbb{Z}$ be a polynomial. The following are equivalent:*

- (i) *$((T_1 \times \cdots \times T_d)^{p(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes d}$.*
- (ii) *Every eigenvalue of $T_1 \times \cdots \times T_d$ is uniform for p.*
- (iii) *For every $1 \leq i \leq d$, if λ_i is either 1 or an eigenvalue of T_i, where at least one λ_i is an eigenvalue, then $\prod_{i=1}^{d} \lambda_i$ is uniform for p.*

Proof. For convenience denote $\mathbf{Y} = (Y, \mathcal{D}, \nu, T) = (X^d, \mathcal{B}^{\otimes d}, \mu^{\otimes d}, T_1 \times \cdots \times T_d)$.

(i) $\Rightarrow$ (ii): Suppose that λ is an eigenvalue of T. Let $f \in L^\infty(\nu)$ be a non-ν-a.e. constant function such that $Tf = \lambda f$. By (i),

$$0 = \mathbb{E}_{n\in\mathbb{Z}^L} T^{p(n)} f = \mathbb{E}_{n\in\mathbb{Z}^L} \lambda^{p(n)} f.$$

Since f is not ν-a.e. constant, $\mathbb{E}_{n\in\mathbb{Z}^L} \lambda^{p(n)} = 0$ and so λ is uniform for p.

(ii) $\Rightarrow$ (i): It suffices to show that for all $f \in L^\infty(\nu)$ with $\int_Y f\, d\nu = 0$, we have that

$$\mathbb{E}_{n\in\mathbb{Z}^L} T^{p(n)} f = 0.$$

By Proposition 2.8, it follows that

$$\mathbb{E}_{n\in\mathbb{Z}^L} T^{p(n)} f = \mathbb{E}_{n\in\mathbb{Z}^L} T^{p(n)} \mathbb{E}(f|Z_{T,T}(\mathbf{Y})).$$

By Lemma 2.7, we can approximate $\mathbb{E}(f|Z_{T,T}(\mathbf{Y}))$ in $L^2(\nu)$ by finite linear combinations of eigenfunctions of T. So, we may assume without loss of generality that $\mathbb{E}(f|Z_{T,T}(\mathbf{Y}))$ itself is an eigenfunction of T and $T\mathbb{E}(f|Z_{T,T}(\mathbf{Y})) = \lambda \mathbb{E}(f|Z_{T,T}(\mathbf{Y}))$. Since λ is uniform for p,

$$\mathbb{E}_{n\in\mathbb{Z}^L} T^{p(n)} \mathbb{E}(f|Z_{T,T}(\mathbf{Y})) = \mathbb{E}_{n\in\mathbb{Z}^L} \lambda^{p(n)} \mathbb{E}(f|Z_{T,T}(\mathbf{Y})) = 0$$

and we are done.

(ii) $\Leftrightarrow$ (iii): This is a direct corollary of Lemma 3.1. $\square$

4. PET induction

This section deals and explains the PET induction scheme, which is one of the main tools that we use in order to study expressions of the form (1) and (2).[21] This technique was introduced by Bergelson (in the now classical [6]) to study multiple averages for essentially distinct polynomials in weakly mixing systems and show the joint ergodicity property in that setting. His method used an induction argument via van der Corput lemma, reformulated in his setting, to reduce the "complexity" of the family of polynomials.

Following this pivotal work of Bergelson, variations of the initial PET induction scheme were used to tackle more general cases, as the one in [9] to deal with multiple, commuting T_i's and "nice" families of polynomials, and in [17] to deal with multiple, commuting, T_i's and "standard" families of multi-variable polynomials, which we actually follow here too.

The idea is the following: one runs the van der Corput lemma (vdC-operation) in some family of integer valued functions-sequences satisfying some special property and gets a family also satisfying the special property but of lower "complexity". This allows one to run an inductive argument and arrive at a base case. In our case the base case is when all the iterates are linear.

Of course, in all the different aforementioned cases, one has to do several technical variations in the method. In this paper for example, an essential detail is that whenever we talk about a polynomial with multiple variables, we always treat the first variable as a special one (see below for more details). Also, to the best of our knowledge, it is the first time that via the vdC-operations, while running (the variation of) the PET induction, we track down the coefficients of the polynomials (see Section 6), which is crucial for our arguments.

[21] For us, PET is an abbreviation for "Polynomial Exhaustion Technique" (PET also stands for "Polynomial Ergodic Theorem").

Definition. For a polynomial $p(n; h_1, \ldots, h_s)\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}$, we denote with $\deg(p)$ *the degree of p with respect to n* (for example, for $s = 1, L = 2$, the degree of $p(n_1, n_2; h_{1,1}, h_{1,2}) = h_{1,1}h_{1,2}n_1^2 + h_{1,1}^5 n_2$ is 2).

For a polynomial $p(n; h_1, \ldots, h_s) = (p_1(n; h_1, \ldots, h_s), \ldots, p_d(n; h_1, \ldots, h_s))\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$, we let $\deg(p) = \max\limits_{1\leq i\leq d} \deg(p_i)$ and we say that p is *essentially constant* if $p(n; h_1, \ldots, h_s)$ is independent of the variable n. We say that the polynomials $p, q\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$ are *essentially distinct* if $p - q$ is not essentially constant, and *essentially equal* otherwise.

Actually, for a tuple $\mathbf{q} = (q_1, \ldots, q_\ell)$ with polynomials $q_1, \ldots, q_\ell\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$, we let $\deg(\mathbf{q}) = \max\limits_{1\leq i\leq \ell} \deg(q_i)$. We say that $\mathbf{q}$ is *non-degenerate* if $q_1, \ldots, q_\ell$ are all not essentially constant, and are pairwise essentially distinct.[22]

Fix a $\mathbb{Z}^d$-system $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$. Let $q_1, \ldots, q_\ell\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$ be polynomials and $g_1, \ldots, g_\ell\colon X \times (\mathbb{Z}^L)^s \to \mathbb{R}$ be functions such that each $g_m(\cdot; h_1, \ldots, h_s)$ is an $L^\infty(\mu)$ function bounded by 1 for all $h_1, \ldots, h_s \in \mathbb{Z}, 1 \leq m \leq \ell$. For convenience, let $\mathbf{q} = (q_1, \ldots, q_\ell)$ and $\mathbf{g} = (g_1, \ldots, g_\ell)$. We call $A = (L, s, \ell, \mathbf{g}, \mathbf{q})$ a *PET-tuple*, and for $\kappa \in \mathbb{N}$ we set

$$S(A, \kappa) := \overline{\mathbb{E}}^{\square}_{h_1,\ldots,h_s\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} \prod_{m=1}^{\ell} T_{q_m(n;h_1,\ldots,h_s)} g_m(x; h_1, \ldots, h_s) \Big\|^{\kappa}_{L^2(\mu)}.$$

We define $\deg(A) = \deg(\mathbf{q})$, and we say that A is *non-degenerate* if $\mathbf{q}$ is non-degenerate. For any $f \in L^\infty(\mu)$, we say that $A = (L, s, \ell, \mathbf{g}, \mathbf{q})$ is *standard* for f if there exists $1 \leq m \leq \ell$ such that $\deg(A) = \deg(q_m)$ and $g_m(x; h_1, \ldots, h_s) = f(x)$. That is, f appears as one of the functions in $\mathbf{g}$, only depending on the first variable, and that the polynomial acting on f is of the highest degree. We say $A = (L, s, \ell, \mathbf{g}, \mathbf{q})$ is *semi-standard* for f if there exists $1 \leq m \leq \ell$ such that $g_m(x; h_1, \ldots, h_s) = f(x)$, which is similar to being standard, but we do not require the polynomial acting on f to be of the highest degree.

For each PET-tuple $A = (L, s, \ell, \mathbf{g}, \mathbf{q})$ and polynomial $q\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$, we define the *vdC-operation*, $\partial_q A$, according to the following three steps:

Step 1: For all $1 \leq m \leq \ell$, let $g'_m = g'_{m+\ell} = g_m$, and $q'_1, \ldots, q'_{2\ell}\colon (\mathbb{Z}^L)^{s+2} \to \mathbb{Z}^d$ be polynomials defined as

$$q'_m(n; h_1, \ldots, h_{s+1}) = \begin{cases} q_m(n; h_1, \ldots, h_s) - q(n; h_1, \ldots, h_s) & , 1 \leq m \leq \ell \\ q_{m-\ell}(n + h_{s+1}; h_1, \ldots, h_s) - q(n; h_1, \ldots, h_s) & , \ell + 1 \leq m \leq 2\ell \end{cases},$$

i.e., we subtract the polynomial q from the first ℓ polynomials and for the second ℓ ones we first shift by h_{s+1} about the first variable and then we subtract q.

Step 2: We remove from $q'_1(n; h_1, \ldots, h_{s+1}), \ldots, q'_{2\ell}(n; h_1, \ldots, h_{s+1})$ the polynomials which are essentially constant and the corresponding terms with those as iterates (this will be justified via the use of the Cauchy-Schwarz inequality and the fact that the functions g_m are bounded), and then put the non-essentially constant ones in groups $J_i = \{q''_{i,1}, \ldots, q''_{i,t_i}\}$, $1 \leq i \leq r$ for some r, $t_i \in \mathbb{N}^*$ such that two polynomials are essentially distinct if and only if they belong to different groups. We now write $q''_{i,j}(n; h_1, \ldots, h_{s+1}) = q''_{i,1}(n; h_1, \ldots, h_{s+1}) + p''_{i,j}(h_1, \ldots, h_{s+1})$ for

[22] The separation between using or not bold characters might look confusing in the beginning, it makes it clearer though when we use both vectors and vectors of vectors of polynomials.

some polynomial $p''_{i,j}$ for all $1 \leq j \leq t_i$, $1 \leq i \leq r$. For convenience, we also relabel $g'_1, \dots, g'_{2\ell}$ accordingly as $g''_{i,j}$ for all $1 \leq j \leq t_i$, $1 \leq i \leq r$.

Step 3: For all $1 \leq i \leq r$, let $q^*_i = q''_{i,1}$ and

$$g^*_i(x; h_1, \dots, h_{s+1}) = g''_{i,1}(x; h_1, \dots, h_{s+1}) \prod_{j=2}^{t_i} T_{p''_{i,j}(h_1,\dots,h_{s+1})} g''_{i,j}(x; h_1, \dots, h_{s+1}).$$

Set $\mathbf{q}^* = (q^*_1, \dots, q^*_r)$, $\mathbf{g}^* = (g^*_1, \dots, g^*_r)$ and let this new PET-tuple be $\partial_q A = (L, s+1, r, \mathbf{g}^*, \mathbf{q}^*)$.[23]

In practice, the polynomial q is some of the initial polynomials $q_1, \dots, q_\ell$. Therefore, if $q = q_t$ for some $1 \leq t \leq \ell$, we write $\partial_t A$ instead of $\partial_{q_t} A$ to lighten the notation.

We will use the previous notation and quantifiers for the vdC-operation from now on.

The following important proposition informs us that, modulo some power and some constant which are unimportant for our purpose, the value of $S(\cdot, \cdot)$ grows by using the vdC-operation described above.

Proposition 4.1. *Let $(X, \mathcal{B}, \mu, (T_g)_{g \in \mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system, $A = (L, s, \ell, \mathbf{g}, \mathbf{q})$ a PET-tuple, and $q \colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$ a polynomial. Then $\partial_q A$ is non-degenerate and $S(A, 2\kappa) \leq 4^\kappa S(\partial_q A, \kappa)$ for every $\kappa \in \mathbb{N}$.*

Proof. Since in Step 2 of the vdC-operation, essentially constant polynomials are removed and polynomials which are essentially the same are grouped together, we have that $\partial_q A$ is non-degenerate.

On the other hand, we have that $S(A, 2\kappa)$ equals to

[23] Here we abuse the notation by writing $\partial_q A$ to denote any of such operations obtained from Step 1 to 3. Strictly speaking, $\partial_q A$ is not uniquely defined as the order of grouping of $q'_1, \dots, q'_{2\ell}$ in Step 2 is ambiguous. However, this is done without loss of generality, since the order does not affect the value of $S(\partial_q A, \cdot)$.

$$
\begin{aligned}
&\overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N} \prod_{m=1}^{\ell} T_{q_m(n;h_1,\dots,h_s)} g_m(x;h_1,\dots,h_s)\Big\|^{2\kappa}_{L^2(\mu)} \\
\leq & 4^\kappa \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_{s+1}\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big|\mathbb{E}_{n\in I_N} \Big\langle \prod_{m=1}^{l} T_{q_m(n;h_1,\dots,h_s)} g_m(x;h_1,\dots,h_s), \qquad \text{(by Lemma 2.2)} \\
& \prod_{m=1}^{l} T_{q_m(n+h_{s+1};h_1,\dots,h_s)} g_m(x;h_1,\dots,h_s) \Big\rangle\Big|^{\kappa} \\
= & 4^\kappa \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_{s+1}\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big|\mathbb{E}_{n\in I_N} \Big\langle \prod_{m=1}^{l} T_{q'_m(n;h_1,\dots,h_s,h_{s+1})} g'_m(x;h_1,\dots,h_s), \qquad \text{(invariance of } \mu) \\
& \prod_{m=1}^{l} T_{q'_{m+l}(n;h_1,\dots,h_s,h_{s+1})} g'_m(x;h_1,\dots,h_s) \Big\rangle\Big|^{\kappa} \qquad \text{(end of Step 1)} \\
\leq & 4^\kappa \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_{s+1}\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N} \prod_{i=1}^{r} T_{q''_{i,1}(n;h_1,\dots,h_{s+1})} \big(g''_{i,1}(x;h_1,\dots,h_{s+1}) \cdot \\
& \prod_{j=2}^{t_i} T_{p''_{i,j}(h_1,\dots,h_{s+1})} g''_{i,j}(x;h_1,\dots,h_{s+1})\big)\Big\|^{\kappa}_{L^2(\mu)} \qquad \text{(Cauchy-Schwarz and Step 2)} \\
= & 4^\kappa \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_{s+1}\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N} \prod_{i=1}^{r} T_{q^*_i(n;h_1,\dots,h_{s+1})} g^*_i(x;h_1,\dots,h_{s+1})\Big\|^{\kappa}_{L^2(\mu)} \qquad \text{(Step 3)},
\end{aligned}
$$

which is $4^\kappa S(\partial_q A, \kappa)$, completing the proof. □

The following theorem shows that when we start with a PET-tuple which is standard for a function, then after finitely many vdC-operations, we arrive at a new PET-tuple of degree 1 which is still standard for the same function. This is useful because by [17, Proposition 3.1], whenever we have an average with linear iterates, we can bound the limsup of the norm of the average by some Host-Kra seminorm of the functions. We caution the reader that in our method, we alternate this standard procedure and instead of deriving to linear iterates for "some functions", we run the PET induction multiple times to arrive at linear iterates isolating "each function" separately.

Theorem 4.2. *Let $(X,\mathcal{B},\mu,(T_g)_{g\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system and $f\in L^\infty(\mu)$. If A is a non-degenerate PET-tuple which is standard for f, then there exist $\rho_1,\dots,\rho_t\in\mathbb{N}^*$, for some $t\in\mathbb{N}$, such that $\partial_{\rho_t}\dots\partial_{\rho_1}A$ is a non-degenerate PET-tuple which is standard for f with $\deg(\partial_{\rho_t}\dots\partial_{\rho_1}A)=1$.*

As an example to demonstrate how the method works, we present some computations for our Example 1.

First part of computations for Example 1: For a $\mathbb{Z}^2$-system $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^2})$ and $f_1, f_2 \in L^\infty(\mu)$, the PET-tuple of Example 1 is

$$A = (1, 0, 2, (f_1, f_2), (p_1, p_2)),$$

where $p_1(n) = (n^2 + n, 0) = (n^2 + n)e_1$, $p_2(n) = (0, n^2) = n^2 e_2$, for $e_1 = (1, 0)$ and $e_2 = (0, 1)$. For $i = 1$ and 2, we explain how to find a sequence of vdC-operations to reduce A into a non-degenerate PET-tuple of degree 1 which is standard for f_i.

We first isolate the function f_1. Setting $e = (1, -1)$, we have $\partial_2 A = (1, 1, 3, (f_1, f_1, f_2), \mathbf{p}_1)$, where the tuple $\mathbf{p}_1$ essentially equals to

$$(n^2 e + ne_1, n^2 e + (2h_1 + 1)ne_1, 2h_1 ne_2)$$

(one term is removed because it is essentially constant and so $\ell = 3$). Then $\partial_3\partial_2 A = (1, 2, 4, (f_1, f_1, f_1, f_1), \mathbf{p}_2)$, where the tuple $\mathbf{p}_2$ essentially equals to
$((n^2 + 2h_1 n)e + (1 - 2h_1)ne_1, (n^2 + 2h_1 n)e + ne_1, (n^2 + 2(h_1 + h_2)n)e + (1 - 2h_1)ne_1, (n^2 + 2(h_1 + h_2)n)e + ne_1)$ (two terms are removed because they are essentially constant and so $\ell = 4$). Finally $\partial_2\partial_3\partial_2 A = (1, 3, 7, (f_1, \dots, f_1), \mathbf{p}_3)$, where the tuple $\mathbf{p}_3$ essentially equals to
$(-2h_1 ne_1, 2h_2 ne - 2h_1 e_1, 2h_2 ne, 2h_3 ne - 2h_1 ne_1, 2h_3 ne, 2(h_2 + h_3)ne - 2h_1 ne_1, 2(h_2 + h_3)ne)$ (one term is removed because it is essentially constant and so $\ell = 7$). We have that $\partial_2\partial_3\partial_2 A$ is non-degenerate and standard for f_1, and $\deg(\partial_2\partial_3\partial_2 A) = 1$.

We continue by isolating f_2. Note that $\partial_1 A = (1, 1, 3, (f_2, f_1, f_2), \mathbf{p}_1)$, where the tuple $\mathbf{p}_1$ essentially equals to

$$(-n^2 e - ne_1, 2h_1 ne_1, -n^2 e - ne_1 + 2h_1 ne_2)$$

(one term is removed for it is essentially constant and so $\ell = 3$). Then $\partial_2\partial_1 A = (1, 2, 4, (f_2, f_2, f_2, f_2), \mathbf{p}_2)$, where the tuple $\mathbf{p}_2$ essentially equals to

$$(-n^2 e - (2h_1 + 1)ne_1, -(n^2 + 2h_1 n)e - ne_1, -(n^2 + 2h_2 n)e - (2h_1 + 1)ne_1, -(n^2 + 2(h_1 + h_2)n)e - ne_1)$$

(two terms are removed because they are essentially constant and so $\ell = 4$). Finally $\partial_1\partial_2\partial_1 A = (1, 3, 7, (f_2, \dots, f_2), \mathbf{p}_3)$, where the tuple $\mathbf{p}_3$ essentially equals to
$(2h_1 ne_2, -2h_2 ne, -2h_2 ne + 2h_1 ne_2, -2h_3 ne, -2h_3 ne + 2h_1 ne_2, -2(h_2 + h_3)ne, -2(h_2 + h_3)ne + 2h_1 ne_2)$ (one term is removed because it is essentially constant and so $\ell = 7$). We have that $\partial_1\partial_2\partial_1 A$ is non-degenerate and standard for f_2, and $\deg(\partial_1\partial_2\partial_1 A) = 1$.

Proof of Theorem 4.2. We follow the ideas of the PET induction in [17] and [21].

If $\deg(A) = 1$, there is nothing to prove. So, we assume that $\deg(A) \geq 2$, $A = (L, s, \ell, \mathbf{g} = (g_1, \dots, g_\ell), \mathbf{q} = (q_1, \dots, q_\ell))$, with $q_i = (q_{i,1}, \dots, q_{i,d})$, $1 \leq i \leq \ell$, where each $q_{i,j}$ is a polynomial from $\mathbb{Z}^{s+1}$ to $\mathbb{Z}$. Recall that $\deg(q_i) = \max_{1\leq j\leq d} \deg(q_{i,j})$. In this proof, we are thinking of $\mathbf{q}$ as an $\ell \times d$ matrix $(q_{i,j})_{1\leq i\leq \ell, 1\leq j\leq d}$ with polynomial entries.

We say that $p, q\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}$ are *equivalent*, and we write that $p \sim q$, if $\deg(p) = \deg(q)$ and $\deg(p - q) < \deg(p)$; otherwise, we write $p \nsim q$. It is not hard to see that "$\sim$" defines an equivalence relation. Suppose that $\deg(\mathbf{q}) \leq D$. We define the *column weight* of the column j to be the vector $w_j(\mathbf{q}) = (w_{1,j}(\mathbf{q}), \dots, w_{D,j}(\mathbf{q}))$, where each $w_{k,j}(\mathbf{q})$ is equal to the number of equivalent classes in $\mathbf{q}$ of degree k in the column j (*i.e.*, among $q_{1,j}, \dots, q_{\ell,j}$). For two column weights $\mathbf{v} = (v_1, \dots, v_D)$ and $\mathbf{v}' = (v'_1, \dots, v'_D)$, we say that $\mathbf{v} < \mathbf{v}'$ if there exists $1 \leq k \leq D$ such that $v_k < v'_k$ and $v_{k'} = v'_{k'}$ for all $k' > k$ (notice that we start comparing them from the last coordinate because this is the one associated to the highest degree). Then, the set of weights and the set of column degrees are well ordered sets. Putting this information about $\mathbf{q}$ in rows, we get the $D \times d$ matrix $w_{\mathbf{q}} = [w_1(\mathbf{q}), \dots, w_d(\mathbf{q})]$ which we call the *subweigth* of $\mathbf{q}$.

Given a matrix M (with polynomial entries), we define its *k-reduction*, denoted by $R_k(M)$, to be the submatrix of M obtained by only considering the rows whose first k elements are 0, after discarding these 0's. For instance, for the matrix

$$M = \begin{pmatrix} 0 & p_1 & p_2 & p_3 \\ p_4 & 0 & 0 & p_5 \\ 0 & 0 & p_6 & p_7 \\ 0 & 0 & 0 & p_8 \end{pmatrix}$$

where $p_1, \dots, p_8$ are non-zero polynomials, its i-reduction for $i = 1, 2, 3, 4$ is $\begin{pmatrix} p_1 & p_2 & p_3 \\ 0 & p_6 & p_7 \\ 0 & 0 & p_8 \end{pmatrix}$, $\begin{pmatrix} p_6 & p_7 \\ 0 & p_8 \end{pmatrix}$, $\begin{pmatrix} p_8 \end{pmatrix}$ and $\emptyset$ respectively. By convention, the 0-reduction $R_0(M)$ is M itself and the k-reduction for $k \geq \ell$ is $\emptyset$.

We now define an order associated to matrices. The *weight* of a matrix $\mathbf{q}$ with polynomial entries, denoted by $W(\mathbf{q})$, is the vector of the matrices $(w(R_0(\mathbf{q})), w(R_1(\mathbf{q})), \dots, w(R_{\ell-1}(\mathbf{q})))$, where ℓ is the number of columns of $\mathbf{q}$. Given two polynomial matrices $\mathbf{q}$ and $\mathbf{q}'$, $\deg(\mathbf{q}), \deg(\mathbf{q}') \leq D$, we say that $W(\mathbf{q}') < W(\mathbf{q})$ if there exist $1 \leq J, K \leq \ell$ such that

$$w_j(R_k(\mathbf{q})) = w_j(R_k(\mathbf{q}')) \text{ for all } j < J \text{ and all } k = 0, \dots, \ell - 1;$$

and

$$w_J(R_k(\mathbf{q})) = w_J(R_k(\mathbf{q}')) \text{ for all } k = 0, \dots, K-1 \text{ and } w_J(R_K(\mathbf{q})) < w_J(R_K(\mathbf{q}')).$$

Under this order, the set of weights of matrices is well-ordered. For a PET-tuple $A = (L, s, \ell, \mathbf{g}, \mathbf{q})$, we define $W(A) = W(\mathbf{q})$ to be the *weight* of A.

Claim: Let A be a non-degenerate PET-tuple which is standard for f with $\deg(A) \geq 2$. There exists $1 \leq \rho \leq \ell$ such that $\partial_\rho A$ is non-degenerate and standard for f with $W(\partial_\rho A) < W(A)$.

We first finish the proof of the theorem assuming that the claim holds. Let A be a non-degenerate PET-tuple which is standard for f and $\deg(A) \geq 2$. After using the claim finitely many steps, the decreasing chain $W(A) > W(\partial_{\rho_1} A) > W(\partial_{\rho_2}\partial_{\rho_1} A) > \dots$ will eventually terminate, so we will end up with a non-degenerate PET-tuple $\partial_{\rho_t} \dots \partial_{\rho_1} A$ which is standard for f, with $\deg(\partial_{\rho_t} \dots \partial_{\rho_1} A) = 1$. This finishes the proof.

So it suffices to prove the claim. Relabeling if necessary, we may assume without loss of generality that $g_1 = f$ and $\deg(q_{1,1}) = \deg(A) \geq 2$. Let $j_0 \in \{0, \dots, \ell\}$ be the smallest integer such that $R_{j_0+1}(\mathbf{q}) = \emptyset$. We choose $1 \leq \rho \leq \ell$ in the following way:

(i) Case that $j_0 = 0$. This case has three sub-cases.

(a) If some $q_{i,1} \not\sim q_{1,1}$, then let ρ be the smallest integer such that $q_{\rho,1} \not\sim q_{1,1}$. In this case, since $q_{\rho,1} \not\sim q_{1,1}$ and A is standard for f, $\partial_\rho A$ is standard for f. Moreover, $w_{D,1}(\partial_\rho A) = w_{D,1}(A) - 1$ and so $W(\partial_\rho A) < W(A)$.

(b) If all $q_{1,1}, \dots, q_{\ell,1}$ are equivalent and there exist $2 \leq i \leq \ell$, $1 \leq j \leq d$ such that $q_{i,j} \not\sim q_{1,j}$, and either $\deg(q_{i,j})$ or $\deg(q_{1,j})$ equals to $\deg(\mathbf{q})$, then let ρ be the smallest integer such that there exists $1 \leq j \leq d$ with $q_{\rho,j} \not\sim q_{1,j}$, and either $\deg(q_{\rho,j})$ or $\deg(q_{1,j})$ equals to $\deg(\mathbf{q})$. In this case, since $q_{\rho,j}$ is not equivalent to $q_{1,j}$, and either $\deg(q_{\rho,j})$ or $\deg(q_{1,j})$ equals to $\deg(\mathbf{q})$, $\partial_\rho A$ is standard for f. Moreover, $w_{D,1}(\partial_\rho A) = 0 < w_{D,1}(A)$ and so $W(\partial_\rho A) < W(A)$.

(c) If all $q_{1,1}, \dots, q_{\ell,1}$ are equivalent, and for all $1 \leq j \leq d$, either $\deg(q_{i,j})$ is $\deg(q_{1,j})$ for all $1 \leq i \leq \ell$ or $\deg(q_{i,j}) < \deg(\mathbf{q})$ for all $1 \leq i \leq \ell$, then let $\rho = \ell + 1$.[24]
In this case, $\deg(\partial_\rho A) < \deg(A)$. Since $\deg(q_{1,1}) \geq 2$, we have that $\deg(q_{1,1}(n, h_1, \dots, h_s) - q_{1,1}(n + h_{s+1}, h_1, \dots, h_s)) = \deg(q_{1,1}) - 1 = \deg(\partial_\rho A) \geq 1$. So $\partial_\rho A$ is standard for f. Moreover, $w_{D,1}(\partial_\rho A) = 0 < w_{D,1}(A)$ and so $W(\partial_\rho A) < W(A)$.

(ii) Case that $j_0 > 0$. Consider the reduction $R_{j_0}(\mathbf{q})$ of the matrix $\mathbf{q}$.

(a) Suppose that an entry of the first column of $R_{j_0}(\mathbf{q})$ (which is of course an entry of the $j_0 + 1$ column of $\mathbf{q}$) is not equivalent to any other entry of the first column of $R_{j_0}(\mathbf{q})$. Among such entries, let ρ be the smallest index such that q_{ρ,j_0+1} has minimal degree.
In this case, we have that $\partial_\rho A$ is standard for f. Moreover,

$$w_{\deg(q_{\rho,j_0+1}),1}(\partial_\rho^{j_0} A) > w_{\deg(q_{\rho,j_0+1}),j_0}(R_{j_0}(\mathbf{q})),$$

where $\partial_\rho^k = \partial_\rho \dots \partial_\rho$ (k times). One can check that this implies that $W(\partial_\rho A) < W(A)$.

(b) Suppose all entries in the first column of $R_{j_0}(\mathbf{q})$ are equivalent. Then let ρ be such that q_{ρ,j_0+1} corresponds to the first entry of the first column of $R_{j_0}(\mathbf{q})$.
In this case, $\partial_\rho A$ is standard for f. Moreover,

$$w_{\deg(q_{\rho,j_0+1}),1}(\partial_\rho^{j_0} A) > w_{\deg(q_{\rho,j_0+1}),j_0}(R_{j_0}(\mathbf{q})).$$

One can check that this fact implies that $W(\partial_\rho A) < W(A)$.

This proves the claim and completes the proof. □

We now provide a proof of Proposition 2.8.

Proof of Proposition 2.8. Let $A = (L, 0, 1, \{f\}, \{p\})$. It suffices to show that $S(A, \kappa) = 0$ for some $\kappa \in \mathbb{N}$, assuming that $\mathbb{E}(f|Z_{\mathbb{Z},\mathbb{Z}}(\mathbf{X})) = 0$. For any $s \in \mathbb{N}^*$ and function $u\colon (\mathbb{Z}^L)^s \to \mathbb{Z}$, let $\Delta u\colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}$ be the function $\Delta u(x_1, \dots, x_{s+1}) = u(x_1 + x_{s+1}, \dots, x_s) - u(x_1, \dots, x_s)$ and $\Delta^k u = (\Delta \circ \dots \circ \Delta)u$ (k times).

If $\deg(p) > 1$, then it is easy to verify that $\partial_1 A = (L, 1, 1, \{f\}, \{\Delta p\})$. By induction, $\partial_1^k A = (L, k, 1, \{f\}, \{\Delta^k p\})$ for all $k < \deg(p)$. By Proposition 4.1, we have that $S(A, 2^K) \leq 4^{2^K-1} S(\partial_1^K A, 1)$, where $K = \deg(p) - 1$. It is easy to see that $\deg(\Delta p) = \deg(p) - 1$,[25] and so $\deg(\Delta^K p) = 1$. We may then assume that $\Delta^K p(n, h_1, \dots, h_K) = c(h_1, \dots, h_K) \cdot n + c'(h_1, \dots, h_K)$ for some polynomials $c(h_1, \dots, h_K) \in \mathbb{Z}^L, c'(h_1, \dots, h_K) \in \mathbb{Z}$ of $h_1, \dots, h_K$ with c not being the constant zero vector. By Theorem 2.3,

$$\overline{\mathbb{E}}_{n \in \mathbb{Z}^L} T_{\Delta^K p(n,h_1,\dots,h_K)} f = T_{c'(h_1,\dots,h_K)} \mathbb{E}(f|\mathcal{I}(G(c(h_1, \dots, h_K)))). \tag{10}$$

[24] We leave it to the interested reader to check that (a), (b) and (c) cover all the possibilities in Case (i).
[25] Recall that “deg” only “sees” the first variable.

If $c(h_1,\dots,h_K)\neq 0$, then

$$\mathcal{I}(G(c(h_1,\dots,h_K)))=Z_{G(c(h_1,\dots,h_K))}\subseteq Z_{\mathbb{Z},G(c(h_1,\dots,h_K))}=Z_{\mathbb{Z},\mathbb{Z}},^{26}$$

where in the last equality we used Lemma 2.4 (iv), since $G(c(h_1,\dots,h_K))$ is a finite index subgroup of $\mathbb{Z}$. By Lemma 2.12, the set of $(h_1,\dots,h_K)\in(\mathbb{Z}^L)^K$ such that $c(h_1,\dots,h_K)=0$ is of upper Banach density 0, so

$$\begin{aligned}S(\partial_1^K A,1)=&\overline{\mathbb{E}}^{\square}_{h_1,\dots,h_K\in\mathbb{Z}^L}\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}}\overline{\lim_{N\to\infty}}\left\|\mathbb{E}_{n\in I_N}T_{\Delta^K p(n,h_1,\dots,h_K)}f\right\|_{L^2(\mu)}\\ =&\overline{\mathbb{E}}^{\square}_{h_1,\dots,h_K\in\mathbb{Z}^L}\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}}\overline{\lim_{N\to\infty}}\left\|\mathbb{E}_{n\in I_N}T_{\Delta^K p(n,h_1,\dots,h_K)}\mathbb{E}(f|Z_{\mathbb{Z},\mathbb{Z}})\right\|_{L^2(\mu)}=0.\end{aligned}$$

This implies that $S(A,2^K)=0$, which finishes the proof. □

5. Characterizing multiple averages along polynomials

In this section we state Theorem 5.1, the stronger form of Theorem 1.1, which is the main contribution of this work. Its validity implies (see below) both Theorems 1.3 and 1.4, our main joint ergodicity results.

5.1. **Characteristic factors for multiple averages.** Recall that a family of (integer valued) polynomials $p_1,\dots,p_k\colon\mathbb{Z}^L\to\mathbb{Z}^d$ is non-degenerate if p_i,p_i-p_j are not essentially constant for all $1\leq i,j\leq k$, $i\neq j$. The following theorem states that in order to study multiple averages along polynomials, it suffices to assume that all the functions f_i are measurable with respect to certain Host-Kra characteristic factors:

Theorem 5.1 (Characteristic factors for multiple averages along polynomials)**.** *Let $d,k,L\in\mathbb{N}^*$ and $p_1,\dots,p_k\colon\mathbb{Z}^L\to\mathbb{Z}^d$ be a non-degenerate family of polynomials of degree at most K. Suppose that $p_i(n)=\sum_{v\in\mathbb{N}^L,|v|\leq K}b_{i,v}n^v$ for some $b_{i,v}\in\mathbb{Q}^d$. Let $R\subseteq\mathbb{Q}^d$ be the set*

$$R\coloneqq\bigcup_{v\in\mathbb{N}^L,0<|v|\leq K}\{b_{i,v},b_{i,v}-b_{i',v}\colon 1\leq i,i'\leq k\}\backslash\{\mathbf{0}\}.$$

Let $(X,\mathcal{B},\mu,(T_g)_{g\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system. For every $f_1,\dots,f_k\in L^\infty(\mu)$, we have that

$$\mathbb{E}_{n\in\mathbb{Z}^L}T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k=0 \text{ if } \mathbb{E}(f_i|Z_{\{G(r)^{\times\infty}\}_{r\in R}})=0 \text{ for some } 1\leq i\leq k. \tag{11}$$

In particular, if $(T_g)_{g\in G(r)}$ is ergodic for μ for all $r\in R$, then for every $f_1,\dots,f_k\in L^\infty(\mu)$,

$$\mathbb{E}_{n\in\mathbb{Z}^L}T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k=0 \text{ if } \mathbb{E}(f_i|Z_{(\mathbb{Z}^d)^{\times\infty}})=0 \text{ for some } 1\leq i\leq k. \tag{12}$$

Remark. The following weaker form of (11) in Theorem 5.1 can be derived by the results of [17]:

$$\mathbb{E}_{n\in\mathbb{Z}^L}T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k=0 \text{ if } \mathbb{E}(f_i|Z_{\{G(r)^{\times\infty}\}_{r\in\mathbb{Z}^d\backslash\{\mathbf{0}\}}})=0 \text{ for some } 1\leq i\leq k.$$

[26] Note that one can not conclude that $\mathcal{I}(G(c(h_1,\dots,h_K)))=Z_{G(c(h_1,\dots,h_K))}=Z_{\mathbb{Z}}$ because Lemma 2.4 (iv) is invalid for $d=1$.

Hence, (12) holds if T_g is assumed to be ergodic for μ for all $g \in \mathbb{Z}^d \backslash \{\mathbf{0}\}$. Theorem 5.1 improves the result of [17] since one only needs to require finitely many T_g's to be ergodic (*i.e.*, the generators of $G(r), r \in R$) in order to deduce (12).

On the other hand, it is worth noting that (11) has room for improvement (meaning that it is possible for one to replace the factor $Z_{\{G(r)^{\times\infty}\}_{r\in R}}$ of (11) with smaller ones), as we shall see in the examples below. Actually, we do have a stronger version of (11) (see the proof of Theorem 5.1), but (11) already captures the essence of our result as it is stated here.

Another important example of polynomial averages is the following, for which we actually characterize its convergence to the "expected" limit, where all the transformations have the same polynomial iterate.

Example 2. Let $(X, \mathcal{B}, \mu, T_1, \dots, T_d)$ be a system with commuting transformations. One should think of $T_1, \dots, T_d$ as a $\mathbb{Z}^d$-action $(S_g)_{g\in\mathbb{Z}^d}$ with $T_i = S_{e_i}$, where we recall that $e_i \in \mathbb{Z}^d$ denotes the vector whose ith entry is 1 and all other entries are 0's. Let $p_1, \dots, p_d \colon \mathbb{Z} \to \mathbb{Z}^d$ be polynomials given by $p_i(n) = p(n)e_i$ for some polynomial $p \colon \mathbb{Z} \to \mathbb{Z}$. By Theorem 5.1, we have that

$$\mathbb{E}_{n\in\mathbb{Z}} T_1^{p(n)} f_1 \cdot \ldots \cdot T_d^{p(n)} f_d = 0 \text{ if } \mathbb{E}(f_i | Z_{\{G(r)^{\times\infty}\}_{r\in R}}) = 0 \text{ for some } 1 \le i \le d, \tag{13}$$

where $R = \{T_i, T_i T_j^{-1} \colon 1 \le i, j \le d, i \neq j\}$. We remark that $Z_{\{G(r)^{\times\infty}\}_{r\in R}}$ is not necessarily the smallest factor with this property. For example, if $p(n) = n$, then (13) is a weaker form of Proposition 6.1 (or [15, Proposition 1]).

Continuation of Example 1. Recall the $\mathbb{Z}^2$-system $\mathbf{X}$ with two commuting transformations T_1, T_2 and $p_1, p_2 \colon \mathbb{Z} \to \mathbb{Z}^2$ polynomials given by $p_1(n) = (n^2 + n, 0)$ and $p_2(n) = (0, n^2)$. By Theorem 5.1, we have that

$$\mathbb{E}_{n\in\mathbb{Z}} T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2 = 0 \text{ if } \mathbb{E}(f_i | Z_{\{G(r)^{\times\infty}\}_{r\in R}}) = 0 \text{ for } i = 1 \text{ or } 2, \tag{14}$$

where $R = \{T_1, T_2, T_1 T_2^{-1}\}$. Again $Z_{\{G(r)^{\times\infty}\}_{r\in R}}$ is not the smallest factor with this property (later, in equality (36), we will obtain an improvement of (14)).

It is an interesting, in general open (and definitely hard), question to ask what are the smallest factors $Z_1, \dots, Z_k$ of $\mathbf{X}$ such that for every $f_1, \dots, f_k \in L^\infty(\mu)$,

$$\mathbb{E}_{n\in\mathbb{Z}} T_{p_1(n)} f_1 \cdot \ldots \cdot T_{p_k(n)} f_k = 0 \text{ if } \mathbb{E}(f_i | Z_i) = 0 \text{ for some } 1 \le i \le k.$$

5.2. **Proofs of the joint ergodicity results assuming Theorem 5.1.** In this subsection we explain how to derive our main joint ergodicity results, Theorems 1.3 and 1.4, assuming the validity of Theorem 5.1.

Proof of Theorem 1.3 assuming Theorem 5.1. Let R be defined as in Theorem 5.1. Since T_g is ergodic for all $g \in R$, by Theorem 5.1, we may assume without loss of generality that all $f_1, \dots, f_k$ are measurable with respect to $Z_{(\mathbb{Z}^d)^{\times\infty}}(\mathbf{X})$ (note that conditions (i) and (ii) remain valid when passing to a factor system). By $L^1(\mu)$-approximation, we may assume without loss of generality that all $f_1, \dots, f_k$ are measurable with respect to $Z_{(\mathbb{Z}^d)^{\times M}}(\mathbf{X})$ for some $M \in \mathbb{N}$. By Theorem 2.6 and again by $L^1(\mu)$-approximation, we may further assume without loss of generality that all $f_1, \dots, f_k$ are measurable with respect a factor of $\mathbf{X}$ which is isomorphic to an $(M-1)$-step $\mathbb{Z}^d$-nilsystem.

So, we may assume that $\mathbf{X}$ is itself an $(M-1)$-step $\mathbb{Z}^d$-nilsystem. Since $(T_{p_1(n)} \times \cdots \times T_{p_k(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes k}$, we have that

$$\mathbb{E}_{n\in\mathbb{Z}^L} T_{p_1(n)} f_1(x_1) \cdot \ldots \cdot T_{p_k(n)} f_k(x_k) = \int_X f_1 \, d\mu \cdot \ldots \cdot \int_X f_k \, d\mu \tag{15}$$

for $\mu^{\otimes k}$-a.e. $(x_1,\ldots,x_k) \in X^k$. By Theorem 2.9, (15) holds for all $(x_1,\ldots,x_k) \in X^k$ and so in particular

$$\mathbb{E}_{n\in\mathbb{Z}^L} T_{p_1(n)} f_1(x) \cdot \ldots \cdot T_{p_k(n)} f_k(x) = \int_X f_1 \, d\mu \cdot \ldots \cdot \int_X f_k \, d\mu \tag{16}$$

for all $x \in X$ (as a pointwise limit). By the dominated convergence theorem, (16) also holds as an $L^2(\mu)$-limit, which finishes the proof. □

Before we proceed with the proof of Theorem 1.4, we need the following lemma and proposition:

Lemma 5.2. *Let $(X,\mathcal{B},\mu,T_1,\ldots,T_d)$ be a system with commuting transformations. Then in the product space $(X^d,\mathcal{B}^d,\mu^{\otimes d})$, the σ-algebra of $T_1\times\cdots\times T_d$-invariant sets is measurable with respect to $\bigotimes_{i=1}^{d} Z_{T_i,T_i}$.*

Proof. It suffices to show that $\mathbb{E}(f_1\otimes\cdots\otimes f_d|\mathcal{I}(T_1\times\cdots\times T_d)) = 0$ whenever $\mathbb{E}(f_i|Z_{T_i,T_i}) = 0$ for some $1 \le i \le d$. Without loss of generality, we may assume that all functions are bounded by 1 in $L^\infty(\mu)$ and that $\mathbb{E}(f_1|Z_{T_1,T_1}) = 0$ (or equivalently $\|f_1\|_{T_1,T_1} = 0$). By Lemma 2.1 and Jensen's inequality, setting $a(n) = T_1^n f_1 \otimes \cdots \otimes T_d^n f_d$, we have that

$$\begin{aligned}
\|\mathbb{E}(f_1\otimes\cdots\otimes f_d|\mathcal{I}(T_1\times\cdots\times T_d))\|^4_{L^2(\mu^{\otimes d})} &= \|\mathbb{E}_{n\in\mathbb{Z}}(T_1\times\cdots\times T_d)^n f_1\otimes\cdots\otimes f_d\|^4_{L^2(\mu^{\otimes d})} \\
&\le \left(4\mathbb{E}^{\square}_{h\in\mathbb{Z}}\mathbb{E}_{n\in\mathbb{Z}}\langle a(n), a(n+h)\rangle\right)^2 \\
&= 16\mathbb{E}^{\square}_{h\in\mathbb{Z}}\mathbb{E}_{n\in\mathbb{Z}}\langle f_1\otimes\cdots\otimes f_d, T_1^h f_1\otimes\cdots\otimes T_d^h f_d\rangle^2 \\
&\le 16\mathbb{E}^{\square}_{h\in\mathbb{Z}}\left|\int f_1\cdot T_1^h f_1 d\mu\right|^2 \\
&\le 16\mathbb{E}^{\square}_{h\in\mathbb{Z}}\left|\int \mathbb{E}(f_1\cdot T_1^h f_1|\mathcal{I}(T_1))d\mu\right|^2 \\
&\le 16\mathbb{E}^{\square}_{h\in\mathbb{Z}}\left\|\mathbb{E}(f_1\cdot T_1^h f_1|\mathcal{I}(T_1))\right\|^2_{L^2(\mu)} \\
&= 16\|f_1\|^4_{T_1,T_1},
\end{aligned}$$

where the last line follows, for instance, from Lemma 2.4 (iii). This finishes the proof. □

Proposition 5.3. *Let $d, L \in \mathbb{N}^*$, $p\colon \mathbb{Z}^L \to \mathbb{Z}$ a polynomial and $(X,\mathcal{B},\mu,T_1,\ldots,T_d)$ a system with commuting transformations such that $(T_1^{p(n)},\ldots,T_d^{p(n)})_{n\in\mathbb{Z}^L}$ is jointly ergodic for μ. Then*

(i) *$((T_iT_j^{-1})^{p(n)})_{n\in\mathbb{Z}^L}$ is ergodic for μ for all $1\le i,j\le d, i\ne j$; and*
(ii) *$(T_1^{p(n)}\times\cdots\times T_d^{p(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes d}$.*

Proof. The idea of the proof for Part (i) is similar to [4, Proposition 2.1]. Since the language we use is different, we present the proof for completeness.

By assumption,

$$\mathbb{E}_{n\in\mathbb{Z}^L} T_1^{p(n)} f_1 \cdot \ldots \cdot T_d^{p(n)} f_d = \int_X f_1\, d\mu \cdot \ldots \cdot \int_X f_d\, d\mu \tag{17}$$

for all $f_1, \ldots, f_d \in L^\infty(\mu)$. Suppose first that (i) fails. We may assume without loss of generality that $((T_1T_2^{-1})^{p(n)})_{n\in\mathbb{Z}^L}$ is not ergodic for μ. So there exist $g \in L^\infty(\mu)$ not μ-a.e. equal to a constant function and a function $g' \in L^2(\mu)$ such that

$$g' := \mathbb{E}_{n\in\mathbb{Z}^L} (T_1T_2^{-1})^{p(n)} g \neq \int_X g\, d\mu$$

(the existence of the limit can be seen by Furstenberg's classical spectral-theorem approach, or even by a known "non-spectral" approach due to Bergelson). Then $\int_X g'\, d\mu = \int_X g\, d\mu$. Note that g' cannot be μ-a.e. equal to a constant. Letting $f_1 = g, f_2 = g'$ and $f_0 = f_3 = \cdots = f_d \equiv 1$, we have that

$$\begin{aligned}
\int_X f_0 \cdot \mathbb{E}_{n\in\mathbb{Z}^L} T_1^{p(n)} f_1 \cdot \ldots \cdot T_d^{p(n)} f_d\, d\mu &= \mathbb{E}_{n\in\mathbb{Z}^L} \int_X T_1^{p(n)} g \cdot T_2^{p(n)} g'\, d\mu \\
&= \mathbb{E}_{n\in\mathbb{Z}^L} \int_X (T_1T_2^{-1})^{p(n)} g \cdot g'\, d\mu \\
&= \int_X \mathbb{E}_{n\in\mathbb{Z}^L} (T_1T_2^{-1})^{p(n)} g \cdot g'\, d\mu \\
&= \int_X g'^2\, d\mu > \Big(\int_X g'\, d\mu\Big)^2 = \Big(\int_X g\, d\mu\Big)^2 \\
&= \int_X f_1\, d\mu \cdot \ldots \cdot \int_X f_d\, d\mu,
\end{aligned}$$

a contradiction to (17), proving (i).

To show (ii), it suffices to show that for all $f_1, \ldots, f_d \in L^\infty(\mu)$ with $\prod_{i=1}^d \int_X f_i\, d\mu = 0$, we have that

$$\mathbb{E}_{n\in\mathbb{Z}^L} \bigotimes_{i=1}^d T_i^{p(n)} f_i = \mathbb{E}_{n\in\mathbb{Z}^L} (T_1 \times \cdots \times T_d)^{p(n)} \bigotimes_{i=1}^d f_i = 0. \tag{18}$$

We first claim that

$$\mathbb{E}_{n\in\mathbb{Z}^L} \bigotimes_{i=1}^d T_i^{p(n)} f_i = 0 \text{ if } \mathbb{E}(f_i | Z_{\mathbb{Z}^d,\mathbb{Z}^d}(\mathbf{X})) = 0 \text{ for some } 1 \leq i \leq d.$$

We apply the proof of Proposition 2.8 to the $\mathbb{Z}$-system $(X^d, \mathcal{B}^d, \mu^{\otimes d}, T_1 \times \cdots \times T_d)$. Suppose that $\mathbb{E}(f_i | Z_{\mathbb{Z}^d,\mathbb{Z}^d}(\mathbf{X})) = 0$ for some $1 \leq i \leq d$. By Theorem 2.3 and (10) in the proof of Proposition 2.8, it suffices to show that the set of $(h_1, \ldots, h_K) \in (\mathbb{Z}^L)^K$ such that

$$\mathbb{E}\Big(\bigotimes_{i=1}^d f_i | \mathcal{I}(G(c(h_1, \ldots, h_K))\Big) = 0 \tag{19}$$

is of density 1, where $c\colon (\mathbb{Z}^L)^K \to \mathbb{Z}$ is a non-constant polynomial. If $c(h_1,\dots,h_K)\neq 0$, then $\mathcal{I}(G(c(h_1,\dots,h_K))$ is the sub-σ-algebra of $\mathcal{B}^d$ consists of the $(T_1\times\dots\times T_d)^{c(h_1,\dots,h_K)}$-invariant sets. By Lemma 5.2,

$$\mathcal{I}(G(c(h_1,\dots,h_K))\subseteq \bigotimes_{i=1}^d Z_{T_i^{c(h_1,\dots,h_K)},T_i^{c(h_1,\dots,h_K)}}=\bigotimes_{i=1}^d Z_{T_i,T_i},$$

where we used Lemma 2.4 (iv) in the last equality. On the other hand, by (17), we have that $(T_i^{p(n)})_{n\in\mathbb{Z}^L}$ is ergodic for μ for all $1\leq i\leq d$, which implies that T_i is ergodic for μ. By Lemma 2.4 (ii), we have that

$$\mathcal{I}(G(c(h_1,\dots,h_K))\subseteq \bigotimes_{i=1}^d Z_{T_i,T_i}=\bigotimes_{i=1}^d Z_{\mathbb{Z}^d,\mathbb{Z}^d}.$$

Since $\mathbb{E}(f_i|Z_{\mathbb{Z}^d,\mathbb{Z}^d}(\mathbf{X}))=0$, we have that $\mathbb{E}\Big(\bigotimes_{i=1}^d f_i\Big|\bigotimes_{i=1}^d Z_{\mathbb{Z}^d,\mathbb{Z}^d}\Big)=\bigotimes_{i=1}^d \mathbb{E}(f_i|Z_{\mathbb{Z}^d,\mathbb{Z}^d})=0$, and so (19) holds whenever $c(h_1,\dots,h_K)\neq 0$. By Proposition 2.8, such tuples $(h_1,\dots,h_K)$ are of density 1. This proves the claim.

By the claim, it now suffices to prove (18) under the assumption that all f_i are measurable with respect to $Z_{\mathbb{Z}^d,\mathbb{Z}^d}$. By Lemma 2.7, we can approximate each f_i in $L^2(\mu)$ by eigenfunctions of $\mathbf{X}$. By multi-linearity, we may assume without loss of generality that each f_i is a non-constant eigenfunction of $\mathbf{X}$ given by $T_gf_i=\lambda_i(g)f_i$ for all $g\in\mathbb{Z}^d$ for some group homomorphism $\lambda_i\colon\mathbb{Z}^d\to\mathbb{S}^1$ and that $f_i(x)\neq 0$ μ-a.e $x\in X$. Then by (17),

$$0=\prod_{i=1}^d\int_X f_i d\mu=\mathbb{E}_{n\in\mathbb{Z}^L}\prod_{i=1}^d T_i^{p(n)}f_i=\Big(\mathbb{E}_{n\in\mathbb{Z}^L}\prod_{i=1}^d\lambda_i(p(n)e_i)\Big)\prod_{i=1}^d f_i.$$

This implies that $\mathbb{E}_{n\in\mathbb{Z}^L}\prod_{i=1}^d\lambda_i(p(n)e_i)=0$. So,

$$\mathbb{E}_{n\in\mathbb{Z}^L}\bigotimes_{i=1}^d T_i^{p(n)}f_i=\Big(\mathbb{E}_{n\in\mathbb{Z}^L}\prod_{i=1}^d\lambda_i(p(n)e_i)\Big)\bigotimes_{i=1}^d f_i=0.$$

This proves (ii) and finishes the proof. □

Proof of Theorem 1.4 assuming Theorem 5.1. We first prove the "if" part. We want to show that

$$\mathbb{E}_{n\in\mathbb{Z}^L}T_1^{p(n)}f_1\cdot\ldots\cdot T_d^{p(n)}f_d=\int_X f_1\,d\mu\cdot\ldots\cdot\int_X f_d\,d\mu \tag{20}$$

for all $f_1,\dots,f_d\in L^\infty(\mu)$.

Regard T_i as T_{e_i} and let $p_1,\dots,p_d\colon\mathbb{Z}^L\to\mathbb{Z}^d$ be polynomials given by $p_i(n)=p(n)e_i$. In this case, there exists $q\in\mathbb{Q}\backslash\{0\}$ such that the set R defined in Theorem 5.1 is $R=\{qe_i,q(e_i-e_j)\colon 1\leq i,j\leq d, i\neq j\}$. By assumption (i), all the $T_iT_j^{-1}$'s (or $T_{e_i-e_j}$'s), $i\neq j$ are ergodic for μ, and so $(T_g)_{g\in G(q(e_i-e_j))}=(T_g)_{g\in G(e_i-e_j)}$ is ergodic for μ. By assumption (ii), $(T_i^{p(n)})_{n\in\mathbb{Z}^L}$ is ergodic for μ for all $1\leq i\leq d$, which implies that T_i (or T_{e_i}) is ergodic for μ. So $(T_g)_{g\in G(qe_i)}=(T_g)_{g\in G(e_i)}$ is ergodic for μ. Thus the assumptions of Theorem 5.1 are fulfilled.

By Theorem 5.1, we may assume without loss of generality that $\mathbf{X} = Z_{(\mathbb{Z}^d)^\times}(\mathbf{X})$ (note that conditions (i) and (ii) remain valid when passing to a factor system). Since $(T_1^{p(n)} \times \dots \times T_d^{p_d(n)})_{n\in\mathbb{Z}^L}$ is ergodic for $\mu^{\otimes d}$, a similar argument, using Theorem 2.9, as in the proof of Theorem 1.3, yields the "if" part of this theorem.

To prove the "only if" part, assume that (20) holds for all $f_1, \dots, f_d \in L^\infty(\mu)$. If $T_iT_j^{-1}$ is not ergodic for some $1 \le i, j \le d, i \ne j$, then there exists $g \in L^\infty(\mu)$ which is not μ-a.e. equal to a constant such that $T_ig = T_jg$. So

$$\mathbb{E}_{n\in\mathbb{Z}^L}(T_iT_j^{-1})^{p(n)}g = \mathbb{E}_{n\in\mathbb{Z}^L} g = g \ne \int_X g d\mu,$$

which implies that $((T_iT_j^{-1})^{p(n)})_{n\in\mathbb{Z}^L}$ is not ergodic for μ, a contradiction to (i) in Proposition 5.3. This proofs (i).

Since (20) holds, (ii) follows directly from the statement (ii) of Proposition 5.3 and the proof is complete. □

5.3. **Ingredients to proving Theorem 5.1.** The rest of the paper is devoted to the proof of Theorem 5.1. In order to keep track of the coefficients of the polynomials after the iterated van der Corput operations, we introduce the following definition:

Definition. Let $d \in \mathbb{N}^*$ and V denote the collection of all finite subsets $\{u_1, \dots, u_k\} \subseteq \mathbb{Q}^d$ containing the zero vector $\mathbf{0}$. For $R_1 = \{u_1, \dots, u_k\} \in V$ and $R_2 \subseteq \mathbb{Q}^d$, we say that R_1 is *equivalent* to R_2 (denoted as $R_1 \sim R_2$) if there exists $1 \le i \le k$ such that $R_2 = \{-ru_i, r(u_j - u_i) \colon 1 \le j \le k\}$ for some $r \in \mathbb{Q}\backslash\{0\}$. Note that $R_1 \sim R_2$ implies that R_1 and R_2 have the same cardinality.[27]

Lemma 5.4. *The relation $\sim$ is an equivalence relation on V.*

Proof. If $R_1 = \{u_1, \dots, u_k\}$ and $u_i = 0$, then $R_1 = \{-ru_i, r(u_j - u_i) \colon 1 \le j \le k\}$ for $r = 1$, and so $R_1 \sim R_1$. Suppose that $R_1 \sim R_2$. We may write $R_1 = \{u_1, \dots, u_k\}$ and $R_2 = \{v_1, \dots, v_k\}$, where $v_i = -ru_i$ and $v_j = r(u_j - u_i)$ for all $1 \le j \le k, j \ne i$ for some $1 \le i \le k$. It follows that $u_i = -(1/r)v_i$ and $u_j = (1/r)(v_j - v_i)$ which means $R_2 \sim R_1$.

Assume now that $R_1 \sim R_2$ and $R_2 \sim R_3$. We may write R_2 as above and $R_3 = \{w_1, \dots, w_k\}$, where $w_{i'} = -r'v_{i'}$ and $w_j = r'(v_j - v_{i'})$ for all $j \ne i'$ for some $1 \le i' \le k$.

If $i = i'$, then $w_i = -r'v_i = -r'(-ru_i) = rr'u_i$, and $w_j = r'(v_j - v_i) = r'r(u_j - u_i) - r'(-ru_i) = rr'u_j$ for all $j \ne i$. So $R_3 = rr'R_1$. This implies that $R_1 \sim R_3$.

If $i \ne i'$, then $w_i = r'(v_i - v_{i'}) = r'(-ru_i) - r'r(u_{i'} - u_i) = -rr'u_{i'}$, $w_{i'} = -r'v_{i'} = r'r(u_i - u_{i'})$, and $w_j = r'(v_j - v_{i'}) = r'r(u_j - u_i) - r'r(u_{i'} - u_i) = rr'(u_j - u_{i'})$ for all $j \ne i, i'$. This implies that $R_1 \sim R_3$ and the result follows. □

We write $R_1 \lesssim R_2$ for some $R_1, R_2 \in V$ if there exists $R_3 \in V$ such that $R_2 \sim R_3$ and $R_1 \subseteq R_3$.

Recall that for $\mathbf{b} = (b_1, \dots, b_L) \in (\mathbb{Q}^d)^L$, $b_i \in \mathbb{Q}^d$, we denote $G(\mathbf{b}) = \text{span}_{\mathbb{Q}}\{b_1, \dots, b_L\} \cap \mathbb{Z}^d$. The first ingredient we need to prove Theorem 5.1 is an upper bound for the multiple averages in terms of Host-Kra seminorms. The following proposition shows that we can somehow control the coefficients we get in the end of the PET-induction by the initial ones.

[27] Note that $\mathbf{0} \in R_2$ as $r(u_j - u_i) = \mathbf{0}$ for $j = i$.

Proposition 5.5 (Bounding multiple averages by averaged Host-Kra seminorms). *Let $d,k,L\in\mathbb{N}^*$, $p_1,\dots,p_k\colon\mathbb{Z}^L\to\mathbb{Z}^d$ a non-degenerate family of polynomials of degrees at most K, with $p_i(n)=\sum_{v\in\mathbb{N}^L,|v|\leq K} b_{i,v}n^v$ for some $b_{i,v}\in\mathbb{Q}^d$, and $R_v:=\{b_{i,v}\colon 1\leq i\leq k\}\cup\{\mathbf{0}\}$. Then there exist $s,t_1,\dots,t_k\in\mathbb{N}^*$ and polynomials $\mathbf{c}_{i,m}\colon(\mathbb{Z}^L)^s\to(\mathbb{Z}^d)^L, 1\leq i\leq k, 1\leq m\leq t_i$ with $\mathbf{c}_{i,m}\not\equiv\mathbf{0}$, such that the following hold*

(i) *(Control of the coefficients) Each $\mathbf{c}_{i,m}$ is of the form*

$$\mathbf{c}_{i,m}(h_1,\dots,h_s)=\sum_{a_1,\dots,a_s\in\mathbb{N}^L} h_1^{a_1}\dots h_s^{a_s}\cdot\mathbf{u}_{i,m}(a_1,\dots,a_s)$$

for some

$$\mathbf{u}_{i,m}(a_1,\dots,a_s)=(u_{i,m,1}(a_1,\dots,a_s),\dots,u_{i,m,L}(a_1,\dots,a_s))\in(\mathbb{Q}^d)^L$$

with all but finitely many terms being zero for each (i,m). In addition, for all $a_1,\dots,a_s\in\mathbb{N}^L$ not all equal to $\mathbf{0}$ and every $1\leq i\leq k, 1\leq r\leq L$, denoting

$$U_{i,r}(a_1,\dots,a_s):=\{u_{i,m,r}(a_1,\dots,a_s)\in\mathbb{Z}^d\colon 1\leq m\leq t_k\}\cup\{\mathbf{0}\},$$

we have that there exists $v\in\mathbb{N}^L, |v|>0$ such that $U_{i,r}(a_1,\dots,a_s)\lesssim R_v$.

(ii) *(Control of the average) For every $\mathbb{Z}^d$-system $\mathbf{X}=(X,\mathcal{B},\mu,(T_g)_{g\in\mathbb{Z}^d})$ and every $f_1,\dots,f_k\in L^\infty(\mu)$ bounded by 1, we have that*

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}}\ \overline{\lim_{N\to\infty}}\Big\|\mathbb{E}_{n\in I_N}\prod_{i=1}^k T_{p_i(n)}f_i\Big\|_{L^2(\mu)}\leq C\cdot\min_{1\leq i\leq k}\overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s\in\mathbb{Z}^L}\|f_i\|_{(G(\mathbf{c}_{i,m}(h_1,\dots,h_s)))_{1\leq m\leq t_i}},\tag{21}$$

where $C>0$ is a constant depending only on $p_1,\dots,p_k$.[28]

The second ingredient we need in order to show Theorem 5.1 (which is the main novelty of this paper) is to estimate the right hand side of (21) using the concatenation theorem.

Proposition 5.6 (Bounding averaged Host-Kra seminorms by a single one). *Let $p_1,\dots,p_k\colon\mathbb{Z}^L\to\mathbb{Z}^d$ be a family of polynomials. Suppose that there exist $s,t_1,\dots,t_k\in\mathbb{N}^*$ and polynomials $\mathbf{c}_{i,m}\colon(\mathbb{Z}^L)^s\to(\mathbb{Z}^d)^L, 1\leq i\leq k, 1\leq m\leq t_i$, with $\mathbf{c}_{i,m}\not\equiv\mathbf{0}$ given by*

$$\mathbf{c}_{i,m}(h_1,\dots,h_s)=\sum_{a_1,\dots,a_s\in\mathbb{N}^L} h_1^{a_1}\dots h_s^{a_s}\cdot\mathbf{u}_{i,m}(a_1,\dots,a_s)\tag{22}$$

for some $\mathbf{u}_{i,m}(a_1,\dots,a_s)\in(\mathbb{Q}^d)^L$ with all but finitely many terms equal to $\mathbf{0}$ for each (i,m) such that the following holds: if for every $\mathbb{Z}^d$-system $(X,\mathcal{B},\mu,(T_g)_{g\in\mathbb{Z}^d})$ and every $f_1,\dots,f_k\in L^\infty(\mu)$ bounded by 1, we have that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}}\ \overline{\lim_{N\to\infty}}\Big\|\mathbb{E}_{n\in I_N}\prod_{i=1}^k T_{p_i(n)}f_i\Big\|_{L^2(\mu)}\leq C\cdot\min_{1\leq i\leq k}\overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s\in\mathbb{Z}^L}\|f_i\|_{(G(\mathbf{c}_{i,m}(h_1,\dots,h_s)))_{1\leq m\leq t_i}},\tag{23}$$

[28] One can in fact show that C depends only on d,k,L and the highest degree of $p_1,\dots,p_k$. We will not go into more detail about this fact, since it is unimportant for our purposes.

where $C > 0$ is a constant depending only on $p_1, \ldots, p_k$, then letting

$$H_{i,m} = span_{\mathbb{Q}}\{G(\mathbf{u}_{i,m}(a_1, \ldots, a_s)) \colon a_1, \ldots, a_s \in \mathbb{N}^L\} \cap \mathbb{Z}^d,$$

we have that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} T_{p_1(n)} f_1 \cdot \ldots \cdot T_{p_k(n)} f_k \Big\|_{L^2(\mu)} = 0 \text{ if } \min_{1\le i\le k} \|f_i\|_{H_{i,1}^{\times\infty},\ldots,H_{i,t_i}^{\times\infty}} = 0. \tag{24}$$

We now use Propositions 5.5 and 5.6 to show Theorem 5.1, and leave the proofs of Propositions 5.5 and 5.6 to Sections 6 and 7 respectively.

Proof of Theorem 5.1 assuming Propositions 5.5 and 5.6. Let the set R be defined as in Theorem 5.1. We can assume without loss of generality that $\mathbb{E}(f_1|Z_{\{G(r)^{\times\infty}\}_{r\in R}}) = 0$. Suppose that $p_i(n) = \sum_{v\in\mathbb{N}^L, |v|\le K} b_{i,v} n^v$ for some $b_{i,v} \in \mathbb{Q}^d$ and denote $R_v = \{b_{i,v} \colon 1 \le i \le k\} \cup \{\mathbf{0}\}$ as in Proposition 5.5. By Proposition 5.5, there exist $s, t_1, \ldots, t_k \in \mathbb{N}^*$ and polynomials $\mathbf{c}_{i,m} \colon (\mathbb{Z}^L)^s \to (\mathbb{Z}^d)^L, 1 \le i \le k, 1 \le m \le t_i$, with $\mathbf{c}_{i,m} \not\equiv \mathbf{0}$ given by

$$\mathbf{c}_{i,m}(h_1, \ldots, h_s) = \sum_{a_1,\ldots,a_s\in\mathbb{N}^L} h_1^{a_1} \ldots h_s^{a_s} \cdot \mathbf{u}_{i,m}(a_1, \ldots, a_s)$$

for some $\mathbf{u}_{i,m}(a_1, \ldots, a_s) \in (\mathbb{Q}^d)^L$ with all but finitely many terms equal to $\mathbf{0}$ for each (i, m) (and satisfying the additional assumptions given by Proposition 5.5), such that (21) holds. Let

$$H_{i,m} = \text{span}_{\mathbb{Q}}\{G(\mathbf{u}_{i,m}(a_1, \ldots, a_s)) \colon a_1, \ldots, a_s \in \mathbb{N}^L\} \cap \mathbb{Z}^d.$$

By Proposition 5.6,

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} T_{p_1(n)} f_1 \cdot \ldots \cdot T_{p_k(n)} f_k \Big\|_{L^2(\mu)} = 0 \text{ if } \min_{1\le i\le k} \|f_i\|_{H_{i,1}^{\times\infty},\ldots,H_{i,t_i}^{\times\infty}} = 0.\text{[29]} \tag{25}$$

On the other hand, by the description of $\mathbf{c}_{i,m}$, writing

$$\mathbf{u}_{i,m}(a_1, \ldots, a_s) = (u_{i,m,1}(a_1, \ldots, a_s), \ldots, u_{i,m,L}(a_1, \ldots, a_s)),$$

each $u_{i,m,j}(a_1, \ldots, a_s)$ belongs to the set $U_{i,r}$, which is contained in a set equivalent to one of $R_v, v \in \mathbb{N}^L, 0 < |v| \le k$. By the definition of R, $u_{i,m,j}(a_1, \ldots, a_s) = qr$ for some $q \in \mathbb{Q}$ and $r \in R$. Since $\mathbf{c}_{i,m} \not\equiv \mathbf{0}$, there exists $q_m r_m \in H_{1,m}\backslash\{\mathbf{0}\}$ for some $q_m \in \mathbb{Q}$ and $r_m \in R$ for all $1 \le m \le t_1$. So $G(r_m)$ is a subgroup of $H_{1,m}$. By Lemma 2.4, we have that

$$Z_{H_{1,1}^{\times\infty},\ldots,H_{1,t_1}^{\times\infty}} \subseteq Z_{G(r_1)^{\times\infty},\ldots,G(r_{t_1})^{\times\infty}} \subseteq Z_{\{G(r)^{\times\infty}\}_{r\in R}}.$$

Since $\mathbb{E}(f_1|Z_{\{G(r)^{\times\infty}\}_{r\in R}}) = 0$, we have that $\mathbb{E}(f_1|Z_{H_{1,1}^{\times\infty},\ldots,H_{1,t_1}^{\times\infty}}) = 0$, meaning that the right hand side of (25) is 0, which implies that (11) equals to 0.

If in addition, $(T_g)_{g\in G(r)}$ is assumed to be ergodic for all $r \in R$, then by Corollary 2.5, we have that $Z_{\{G(r)^{\times\infty}\}_{r\in R}} = Z_{(\mathbb{Z}^d)^{\times\infty}}$ and the proof is complete. □

[29] Note that we have in fact proved the following stronger version of (11):

$$\mathbb{E}_{n\in\mathbb{Z}^L} T_{p_1(n)} f_1 \cdot \ldots \cdot T_{p_k(n)} f_k = 0 \text{ if } \mathbb{E}(f_i|Z_{\{H_{i,j}\}_{1\le j\le t_i}^{\times\infty}}) = 0 \text{ for some } 1 \le i \le k.$$

6. Proof of Proposition 5.5

Our strategy to show (21) in Proposition 5.5 is the following: We first fix the functions f_i on the right hand side of (21). By a "dimension-increment" argument (see Proposition 6.3 below), for a fixed i, we may assume that p_i has the highest degree among $p_1, \dots, p_k$, making the PET-tuple to be standard for f_i. Then, Theorem 4.2 allows us to control the left hand side of (21) by a PET-tuple of degree 1 which is also standard for f_i. Finally, a Host-Kra-type inequality for linear polynomials (see Proposition 6.1) implies that (21) holds for some polynomials $\mathbf{c}_{i,m}$. Up to this point, the method we use is similar to the one used in [17] and [21] (the main difference is that we have a more explicit upper bound for $\overline{\lim}_{N\to\infty}\left\|\mathbb{E}_{n\in I_N}T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k\right\|_{L^2(\mu)}$ in Proposition 5.5). Our innovation is that in order for the equation (21) to be useful for our purposes, we need a better description of the functions $\mathbf{c}_{i,m}$, which is the content of part (i) of Proposition 5.5.

We start with the linear case of Proposition 5.5 (the special case $L = 1$ was first proved in [15, Proposition 1]):

Proposition 6.1 (Host-Kra inequality for linear $\mathbb{Z}^L$-averages)**.** *Let $d, k, L \in \mathbb{N}^*$, $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ a $\mathbb{Z}^d$-system and $p_1, \dots, p_k\colon \mathbb{Z}^L \to \mathbb{Z}^d$ essentially distinct and essentially non-constant polynomials of degree 1. Suppose that $p_i(n) = \mathbf{u}_i \cdot n + v_i$ for some $\mathbf{u}_i \in (\mathbb{Z}^d)^L, v_i \in \mathbb{Z}^d$ for all $1 \le i \le k$.*[30] *Then for every $f_1, \dots, f_k \in L^\infty(\mu)$ bounded by 1, we have that*

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}}\left\|\mathbb{E}_{n\in I_N}T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k\right\|_{L^2(\mu)} \le C\cdot \min_{1\le i\le k}\|f_i\|_{G(-\mathbf{u}_i),(G(\mathbf{u}_j-\mathbf{u}_i))_{1\le j\le k, j\neq i}},$$

where C is a constant only depending on k. Moreover, writing $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,L}), u_{i,j} \in \mathbb{Z}^d$ and $R_{e_r} = \{u_{i,r}\colon 1 \le i \le k\} \cup \{\mathbf{0}\}$, the set

$$U_{i,r}(\emptyset) = \{-u_{i,r}, u_{j,r} - u_{i,r}\colon 1 \le j \le k\}$$

is equivalent to R_{e_r}.[31]

Proof. We prove the proposition by induction on k. When $k = 1$, let $p_1(n_1, \dots, n_L) = u_{1,1}n_1 + \dots + u_{1,L}n_L + v_1$ for some $u_{1,1}, \dots, u_{1,L}, v_1 \in \mathbb{Z}^d$. By repeatedly using Theorem 2.3, and the fact that the limit exists, we have that

$$\begin{aligned}\mathbb{E}_{n\in\mathbb{Z}^L}T_{p_1(n)}f_1 &= \mathbb{E}_{n_L\in\mathbb{Z}}\dots\mathbb{E}_{n_1\in\mathbb{Z}}T_{u_{1,1}n_1+\dots+u_{1,L}n_L+v_1}f_1\\ &= \mathbb{E}(T_{v_1}f_1|I(u_{1,1})\cap\dots\cap I(u_{1,L})) = \mathbb{E}(T_{v_1}f_1|I(G(\mathbf{u}_1))),\end{aligned}$$

hence, $\|\mathbb{E}_{n\in\mathbb{Z}^L}T_{p_1(n)}f_1\|_{L^2(\mu)} = \|\mathbb{E}(T_{v_1}f_1|I(G(\mathbf{u}_1)))\|_{L^2(\mu)} = \|\mathbb{E}(f_1|I(G(\mathbf{u}_1)))\|_{L^2(\mu)} = \|f_1\|_{G(\mathbf{u}_1)}$.

[30] Here for $\mathbf{u} = (u_1, \dots, u_L) \in (\mathbb{Z}^d)^L$ and $n = (n_1, \dots, n_L) \in \mathbb{N}^L$, $\mathbf{u}\cdot n$ denotes $n_1u_1 + \dots + n_Lu_L \in \mathbb{Z}^d$.

[31] It is not hard to verify that the sets R_{e_r} and $U_{i,r}(\emptyset)$ coincide with the sets R_v and $U_{i,r}(a_1, \dots, a_s)$ defined in Proposition 5.5 (in this case for $s = 0$) respectively. We leave the verification to the interested reader.

Now suppose that the conclusion holds for $k-1$ for some $k \geq 2$. Then by Lemma 2.2, the Cauchy-Schwarz inequality, and using the fact that $p_1, \dots, p_k$ are of degree 1, we have that

$$\begin{aligned}
&\sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} T_{p_1(n)} f_1 \cdot \ldots \cdot T_{p_k(n)} f_k \Big\|_{L^2(\mu)}^{2^k} \\
&\leq 4^{2^{k-1}} \overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big| \mathbb{E}_{n\in I_N} \int_X \prod_{i=1}^k T_{p_i(n)} f_i \cdot \prod_{i=1}^k T_{p_i(n+h)} f_i \, d\mu \Big|^{2^{k-1}} \\
&= 4^{2^{k-1}} \overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big| \int_X T_{p_k(h)} f_k \cdot \mathbb{E}_{n\in I_N} \prod_{i=1}^{k-1} T_{p_i(n)-p_k(n)} (f_i \cdot T_{p_i(h)} f_i) \, d\mu \Big|^{2^{k-1}} \\
&\leq 4^{2^{k-1}} \overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} \prod_{i=1}^{k-1} T_{p_i(n)-p_k(n)} (f_i \cdot T_{p_i(h)} f_i) \Big\|_{L^2(\mu)}^{2^{k-1}}.
\end{aligned}$$

Since $p_i(n) - p_k(n) = \mathbf{u}'_i \cdot n + (v_i - v_k)$, where $\mathbf{u}'_i := \mathbf{u}_i - \mathbf{u}_k$, by induction hypothesis, there is a constant C' depending on $k-1$ such that

$$\begin{aligned}
&4^{2^{k-1}} \overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} \prod_{i=1}^{k-1} T_{p_i(n)-p_k(n)} (f_i \cdot T_{p_i(h)} f_i) \Big\|_{L^2(\mu)}^{2^{k-1}} \\
&\leq 4^{2^{k-1}} C' \cdot \overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \Big\| f_1 \cdot T_{p_1(h)} f_1 \Big\|_{G(-\mathbf{u}'_1), \{G(\mathbf{u}'_1 - \mathbf{u}'_j)\}_{2\leq j\leq k-1}}^{2^{k-1}} \\
&= C \cdot \overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \Big\| f_1 \cdot T_{p_1(h)} f_1 \Big\|_{G(\mathbf{u}_k - \mathbf{u}_1), \{G(\mathbf{u}_1 - \mathbf{u}_j)\}_{2\leq j\leq k-1}}^{2^{k-1}} \\
&= C \cdot \overline{\mathbb{E}}^{\square}_{h\in\mathbb{Z}^L} \Big\| f_1 \cdot T_{p_1(h)} f_1 \Big\|_{\{G(\mathbf{u}_1 - \mathbf{u}_j)\}_{2\leq j\leq k}}^{2^{k-1}} = C \cdot \|f_1\|_{G(-\mathbf{u}_1), \{G(\mathbf{u}_j - \mathbf{u}_1)\}_{2\leq j\leq k}}^{2^k},
\end{aligned}$$

where $C = 4^{2^{k-1}} C'$ and we used Lemma 2.4 (iii) in the last equality. It is clear that the constant C depends only on k. By symmetry,

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} T_{p_1(n)} f_1 \cdot \ldots \cdot T_{p_k(n)} f_k \Big\|_{L^2(\mu)} \leq C \cdot \min_{1\leq i\leq k} \|f_i\|_{G(-\mathbf{u}_i), \{G(\mathbf{u}_j - \mathbf{u}_i)\}_{1\leq j\leq k, j\neq i}}$$

and the claim follows. □

Before proving the general case of Proposition 5.5, we continue with some additional computations for our Example 1.

Second part of computations for Example 1: Recall that we are dealing with the case $(T_1^{n^2+n}, T_2^{n^2})$, with the PET-tuple

$$A = (1, 0, 2, (f_1, f_2), (p_1, p_2)),$$

where $p_1(n) = (n^2+n, 0) = (n^2+n)e_1, p_2(n) = (0, n^2) = n^2 e_2$, and $e_1 = (1,0), e_2 = (0,1)$ and $e = e_1 - e_2$. In this case, $L = 1$ and $d = 2$, $R_1 = \{e_1, \mathbf{0}\}$, $R_2 = \{e_1, e_2, \mathbf{0}\}$ and $R_v = \{\mathbf{0}\}$ for all $v > 2$. Take $s = 3$.

By the first part of computations of Example 1, isolating f_1, we have that $\partial_2\partial_3\partial_2 A = (3, 7, (f_1, \dots, f_1), \mathbf{p}_3)$, where the tuple $\mathbf{p}_3 = (q_1, \dots, q_7)$ essentially equals to

$$(-2h_1ne_1, 2h_2ne - 2h_1e_1, 2h_2ne, 2h_3ne - 2h_1ne_1, 2h_3ne, 2(h_2+h_3)ne - 2h_1ne_1, 2(h_2+h_3)ne).$$

By Propositions 6.1, 4.1 and Lemma 2.4 (iv) and the fact that Host-Kra seminorms are T_g-invariant, we have that

(26)

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| \mathbb{E}_{n\in I_N} T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2 \right\|_{L^2(\mu)}^8 = S(A, 2^3) \leq C \cdot S(\partial_2\partial_3\partial_2 A, 1)$$

$$\leq C \cdot \overline{\mathbb{E}}^{\square}_{h_1,h_2,h_3\in\mathbb{Z}} \|f_1\|_{G(\mathbf{c}_{1,1}(h_1,h_2,h_3)),\dots,G(\mathbf{c}_{1,7}(h_1,h_2,h_3))},$$

where $\mathbf{c}_{1,1}(h_1, h_2, h_3) = -2h_1e_1$, $\mathbf{c}_{1,2}(h_1, h_2, h_3) = 2h_2e$, $\mathbf{c}_{1,3}(h_1, h_2, h_3) = -2h_1e_1 + 2h_2e$, $\mathbf{c}_{1,4}(h_1, h_2, h_3) = 2h_3e$, $\mathbf{c}_{1,5}(h_1, h_2, h_3) = -2h_1e_1+2h_3e$, $\mathbf{c}_{1,6}(h_1, h_2, h_3) = 2(h_2+h_3)e$, $\mathbf{c}_{1,7}(h_1, h_2, h_3) = -2h_1e_1 + 2(h_2 + h_3)e$. This verifies part (ii) of Proposition 5.5 for $i = 1$. Moreover, using the notation in Proposition 5.5, we have that

$$\begin{aligned} U_{1,1}(1,0,0) &= \{-2e_1, \mathbf{0}, -2e_1, \mathbf{0}, -2e_1, \mathbf{0}, -2e_1\} = \{-2e_1, \mathbf{0}\} \sim R_1, \\ U_{1,1}(0,1,0) &= \{\mathbf{0}, 2e, 2e, \mathbf{0}, \mathbf{0}, 2e, 2e\} = \{2e, \mathbf{0}\} \subseteq \{2e, -2e_2, \mathbf{0}\} \sim R_2, \\ U_{1,1}(0,0,1) &= \{\mathbf{0}, \mathbf{0}, \mathbf{0}, 2e, 2e, 2e, 2e\} = \{2e, \mathbf{0}\} \subseteq \{2e, -2e_2, \mathbf{0}\} \sim R_2. \end{aligned}$$

This verifies part (i) of Proposition 5.5 for $i = 1$.

Similarly, by isolating f_2, we have that $\partial_1\partial_2\partial_1 A = (3, 7, (f_2, \dots, f_2), \mathbf{p}_3)$, where the tuple $\mathbf{p}_3$ essentially equals to $(2h_1ne_2, -2h_2ne, -2h_2ne + 2h_1ne_2, -2h_3ne, -2h_3ne + 2h_1ne_2, -2(h_2 + h_3)ne, -2(h_2 + h_3)ne + 2h_1ne_2)$. Analogously to (26), we have

(27)

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| \mathbb{E}_{n\in I_N} T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2 \right\|_{L^2(\mu)}^8 = S(A, 2^3) \leq C \cdot S(\partial_1\partial_2\partial_1 A, 1)$$

$$\leq C \cdot \overline{\mathbb{E}}^{\square}_{h_1,h_2,h_3\in\mathbb{Z}} \|f_2\|_{G(\mathbf{c}_{2,1}(h_1,h_2,h_3)),\dots,G(\mathbf{c}_{2,7}(h_1,h_2,h_3))},$$

where $\mathbf{c}_{2,1}(h_1, h_2, h_3) = 2h_1e_2$, $\mathbf{c}_{2,2}(h_1, h_2, h_3) = -2h_2e + 2h_1e_2$, $\mathbf{c}_{2,3}(h_1, h_2, h_3) = -2h_2e$, $\mathbf{c}_{2,4}(h_1, h_2, h_3) = -2h_3e+2h_1e_2$, $\mathbf{c}_{2,5}(h_1, h_2, h_3) = -2h_3e$, $\mathbf{c}_{2,6}(h_1, h_2, h_3) = -2(h_2+h_3)e+2h_1e_2$, $\mathbf{c}_{2,7}(h_1, h_2, h_3) = -2(h_2 + h_3)e$. This verifies part (ii) of Proposition 5.5 for $i = 2$. Using the notation in Proposition 5.5, we have that

$$\begin{aligned} U_{2,1}(1,0,0) &= \{2e_2, 2e_2, \mathbf{0}, 2e_2, \mathbf{0}, 2e_2, \mathbf{0}\} = \{2e_2, \mathbf{0}\} \subseteq \{-2e, 2e_2, \mathbf{0}\} \sim R_2, \\ U_{2,1}(0,1,0) &= \{\mathbf{0}, -2e, -2e, \mathbf{0}, \mathbf{0}, -2e, -2e\} = \{-2e, \mathbf{0}\} \subseteq \{-2e, 2e_2, \mathbf{0}\} \sim R_2, \\ U_{2,1}(0,0,1) &= \{\mathbf{0}, \mathbf{0}, \mathbf{0}, -2e, -2e, -2e, -2e\} = \{-2e, \mathbf{0}\} \subseteq \{-2e, 2e_2, \mathbf{0}\} \sim R_2. \end{aligned}$$

This verifies part (i) of Proposition 5.5 for $i = 2$.

We now introduce some additional notation that we will use in the general case. Let $d, \ell, L \in \mathbb{N}^*$, $s \in \mathbb{N}$ and $q_1, \dots, q_\ell \colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$ be polynomials. Denote $\mathbf{q} = (q_1, \dots, q_\ell)$, where

$$q_i(n; h_1, \dots, h_s) = \sum_{b,a_1,\dots,a_s\in\mathbb{N}^L} h_1^{a_1} \dots h_s^{a_s} n^b \cdot u_i(b; a_1, \dots, a_s)$$

for some $u_i(b; a_1, \dots, a_s) \in \mathbb{Q}^d$ with all but finitely many being $\mathbf{0}$ for each $1 \le i \le \ell$. For all $b, a_1, \dots, a_s \in \mathbb{N}^L$, denote

$$R_{\mathbf{q}}(b; a_1, \dots, a_s) := \{u_i(b; a_1, \dots, a_s) \colon 1 \le i \le \ell\} \cup \{\mathbf{0}\} \subseteq \mathbb{Q}^d.$$

Roughly speaking, $R_{\mathbf{q}}(b; a_1, \dots, a_s)$ records the coefficients of $\mathbf{q}$ at "level"-$(b; a_1, \dots, a_s)$ (together with the zero vector $\mathbf{0}$).

The following proposition shows that, during the PET-induction process, after applying the vdC-operation to our expression, we can still keep track of the coefficients of the polynomials.

Proposition 6.2 (vdC-operations treat the sets $R_{\mathbf{q}}(b; a_1, \dots, a_s)$ nicely)**.** *Let $d, \ell, L \in \mathbb{N}^*$, $s \in \mathbb{N}$, $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ a $\mathbb{Z}^d$-system, $q_1, \dots, q_\ell \colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$ polynomials and $\mathbf{q} = (q_1, \dots, q_\ell)$. If $A = (L, s, \ell, \mathbf{q})$ with $\partial_w A = (L, s+1, \ell^*, \mathbf{q}^*)$ for some $\ell^* \in \mathbb{N}^*, 1 \le w \le \ell$, where $\mathbf{q}^* = (q_1^*, \dots, q_{\ell^*}^*)$ for some polynomials $q_1^*, \dots, q_{\ell^*}^* \colon (\mathbb{Z}^L)^{s+2} \to \mathbb{Z}^d$, then for all $b, a_1, \dots, a_{s+1} \in \mathbb{N}^L$ not all equal to $\mathbf{0}$, we have that*

$$R_{\mathbf{q}^*}(b; a_1, \dots, a_{s+1}) \lesssim R_{\mathbf{q}}(b + a_{s+1}; a_1, \dots, a_s). \tag{28}$$

Proof. For convenience we write $\mathbf{q}^* \approx (p_1, \dots, p_{\ell'})$ for some polynomials $p_1, \dots, p_{\ell'}$ if $\mathbf{q}^*$ can be obtained by removing all the essential constant polynomials from $p_1, \dots, p_{\ell'}$, ordering the rest into groups such that two polynomials are essentially distinct if and only if they are in different groups, and then picking one polynomial from each group. It is not hard to see that if $\mathbf{q} \approx \mathbf{q}'$, then $R_{\mathbf{q}}(b; a_1, \dots, a_{s+1}) = R_{\mathbf{q}'}(b; a_1, \dots, a_{s+1})$ for all $b, a_1, \dots, a_{s+1} \in \mathbb{N}^L$ not all equal to $\mathbf{0}$.

Denote $q_i' \colon (\mathbb{Z}^L)^{s+2} \to \mathbb{Z}^d$, $q_i'(n; h_1, \dots, h_{s+1}) = q_i(n + h_{s+1}; h_1, \dots, h_s)$ for all $1 \le i \le \ell$. It suffices to show the statement for $\mathbf{q}^* \approx (q_1' - q_1, q_i - q_1, q_i' - q_1 : i \neq 1)$ as the general case follows similarly.

Suppose that

$$q_i(n; h_1, \dots, h_s) = \sum_{b, a_1, \dots, a_s \in \mathbb{N}^L} h_1^{a_1} \dots h_s^{a_s} n^b \cdot u_i(b; a_1, \dots, a_s)$$

for all $1 \le i \le \ell$. Then, one can immediately check that

$$q_i'(n; h_1, \dots, h_{s+1}) = \sum_{b, a_1, \dots, a_{s+1} \in \mathbb{N}^L} h_1^{a_1} \dots h_{s+1}^{a_{s+1}} n^b \cdot \binom{b + a_{s+1}}{b} u_i(b + a_{s+1}; a_1, \dots, a_s).$$ [32]

If $a_{s+1} = 0$, then the coefficient of $h_1^{a_1} \dots h_s^{a_s} n^b$ for $q_1' - q_1$ is $\mathbf{0}$, and for both $q_i - q_1$ and $q_i' - q_1$ are $u_i(b; a_1, \dots, a_s) - u_1(b; a_1, \dots, a_s)$. This implies that $R_{\mathbf{q}^*}(b; a_1, \dots, a_s, 0) = R_{\mathbf{q}'}(b; a_1, \dots, a_s, 0) \lesssim R_{\mathbf{q}}(b; a_1, \dots, a_s)$,[33] which proves (28).

If $a_{s+1} > 0$, then the coefficient of $h_1^{a_1} \dots h_{s+1}^{a_{s+1}} n^b$ for $q_1' - q_1$, $q_i - q_1$ and $q_i' - q_1$ are $\binom{b + a_{s+1}}{b} u_1(b + a_{s+1}; a_1, \dots, a_s)$, $\mathbf{0}$ and $\binom{b + a_{s+1}}{b} u_i(b + a_{s+1}; a_1, \dots, a_s)$ respectively. In this case $R_{\mathbf{q}^*}(b; a_1, \dots, a_{s+1}) = R_{\mathbf{q}'}(b; a_1, \dots, a_{s+1}) \sim R_{\mathbf{q}}(b + a_{s+1}; a_1, \dots, a_s)$, which finishes the proof. $\square$

[32] For $a = (a_1, \dots, a_L), b = (b_1, \dots, b_L) \in \mathbb{N}^L$, $\binom{a}{b}$ denotes the quantity $\prod_{m=1}^{L} \binom{a_m}{b_m}$.

[33] Note that $R_{\mathbf{q}'}(b; a_1, \dots, a_s, 0) \sim R_{\mathbf{q}}(b; a_1, \dots, a_s)$ if and only if one of $u_i(b; a_1, \dots, a_s)$ is $\mathbf{0}$.

Let A be a PET-tuple and $f \in L^\infty(\mu)$. If A is semi-standard but not standard for f, then the PET-induction does not work well enough to provide an upper bound for $S(A,\kappa)$ in terms of the Host-Kra seminorms of f. To overcome this difficulty, we use a "dimension-increment" argument to change A into a new PET-tuple which is standard for f, but at the cost of increasing the dimension from L to $2L$.[34] In fact, this is the main reason that justifies the multi-variable nature of the results in this article.

This "dimension-increment" argument is carried out in the following proposition. The idea essentially comes from [17, 21], but again some additional work needs to be done in order to keep track of the set $R_{\mathbf{q}}(b; a_1, \dots, a_s)$:

Proposition 6.3 (Dimension-increasing property)**.** *Let $L, d, \ell \in \mathbb{N}^*$, $s \in \mathbb{N}$, $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ a $\mathbb{Z}^d$-system, $f \in L^\infty(\mu)$, $q_1, \dots, q_\ell \colon (\mathbb{Z}^L)^{s+1} \to \mathbb{Z}^d$ polynomials, $g_1, \dots, g_\ell \colon X \times (\mathbb{Z}^L)^s \to \mathbb{R}$ functions such that each $g_i(\cdot; h_1, \dots, h_s)$ is an $L^\infty(\mu)$ function bounded by 1 for all $h_1, \dots, h_s \in \mathbb{Z}^L, 1 \leq i \leq \ell$, and let $\mathbf{q} = (q_1, \dots, q_\ell)$ and $\mathbf{g} = (g_1, \dots, g_\ell)$.*

If the PET-tuple $A = (L, s, \ell, \mathbf{g}, \mathbf{q})$ is non-degenerate and semi-standard but not standard for f, then there exist polynomials $q'_1, \dots, q'_{2\ell-1} \colon (\mathbb{Z}^{2L})^{s+1} \to \mathbb{Z}^d$, functions $g'_1, \dots, g'_{2\ell-1} \colon X \times (\mathbb{Z}^{2L})^s \to \mathbb{R}$ such that each $g'_i(\cdot; h_1, \dots, h_s)$ is an $L^\infty(\mu)$ function bounded by 1 for all $h_1, \dots, h_s \in \mathbb{Z}^L, 1 \leq i \leq 2\ell-1$, $\mathbf{q}' = (q'_1, \dots, q'_{2\ell-1})$ and $\mathbf{g}' = (g'_1, \dots, g'_{2\ell-1})$, such that the PET-tuple $A' = (2L, s, 2\ell-1, \mathbf{g}', \mathbf{q}')$ is non-degenerate and standard for f and $S(A, 2\kappa) \leq S(A', \kappa)$ for all $\kappa > 0$. Moreover, for all $b, b', a_1, \dots, a_s, a'_1, \dots, a'_s \in \mathbb{N}^L$ not all equal to $\mathbf{0}$, there exist $b'', a''_1, \dots, a''_s \in \mathbb{N}^L$ not all equal to $\mathbf{0}$, such that

$$R_{\mathbf{q}'}(b, b'; a_1, \dots, a_s, a'_1, \dots, a'_s) \sim R_{\mathbf{q}}(b''; a''_1, \dots, a''_s). \tag{29}$$

Proof. Since A is semi-standard but not standard for f, we may assume without loss of generality that $g_1(x; h_1, \dots, h_s) = f(x)$, $\deg(q_1) < \deg(A)$, and $\deg(q_\ell) = \deg(A)$. For convenience denote $\mathbf{h} = (h_1, \dots, h_s)$ and $\mathbf{h}' = (h'_1, \dots, h'_s)$. For $1 \leq m \leq \ell$ we set

$$q'_m((n, n'); (\mathbf{h}, \mathbf{h}')) := q_m(n; \mathbf{h}) - q_\ell(n'; \mathbf{h}),^{35} \quad \text{and} \quad g'_m(x; (\mathbf{h}, \mathbf{h}')) := g_m(x; \mathbf{h}),$$

while for $1 \leq m \leq \ell - 1$ we set

$$q'_{m+\ell}((n, n'); (\mathbf{h}, \mathbf{h}')) := q_m(n'; \mathbf{h}) - q_\ell(n'; \mathbf{h}), \quad \text{and} \quad g'_{m+\ell}(x; (\mathbf{h}, \mathbf{h}')) := g_m(x; \mathbf{h}).$$

Also, let $\mathbf{q}' = (q'_1, \dots, q'_{2\ell-1})$, $\mathbf{g}' = (g'_1, \dots, g'_{2\ell-1})$ and $A' = (2L, s, 2\ell-1, \mathbf{q}', \mathbf{g}')$.

Since $\deg(q_\ell) = \deg(A)$ and $\deg(q_1) < \deg(A)$, we have that $\deg(q'_1) = \deg(A')$ and moreover $g'_1 = f$. So A' is standard for f. On the other hand, since A is non-degenerate, one can easily see that $q'_1, \dots, q'_{2\ell-1}$ are essentially distinct (note that $q_\ell(n; \mathbf{h}) - q_\ell(n'; \mathbf{h})$ is essentially non-constant). So A' is non-degenerate.

[34] In the papers [17, 21], where similar methods were used, the dimension was increased from L to $3L$ instead.

[35] The notion $(\mathbf{h}, \mathbf{h}')$ refers to the vector $((h_1, h'_1), \dots, (h_s, h'_s)) \in (\mathbb{Z}^{2L})^s$, which we use to simplify the notation.

Recall that $\overline{\mathbb{E}}^{\square}_{\mathbf{h}\in(\mathbb{Z}^L)^s} = \overline{\mathbb{E}}^{\square}_{h_1\in\mathbb{Z}^L}\dots\overline{\mathbb{E}}^{\square}_{h_s\in\mathbb{Z}^L}$. By the fact that the action is measure preserving and the Cauchy-Schwarz inequality, we have that

$$S(A,2\kappa) = \overline{\mathbb{E}}^{\square}_{\mathbf{h}\in(\mathbb{Z}^L)^s} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} \prod_{m=1}^{\ell} T_{q_m(n;\mathbf{h})} g_m(x;\mathbf{h}) \Big\|^{2\kappa}_{L^2(\mu)}$$

$$= \overline{\mathbb{E}}^{\square}_{\mathbf{h}\in(\mathbb{Z}^L)^s} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big| \mathbb{E}_{n,n'\in I_N} \int_X \prod_{m=1}^{\ell} T_{q_m(n;\mathbf{h})} g_m(x;\mathbf{h}) \cdot T_{q_m(n';\mathbf{h})} g_m(x;\mathbf{h})\, d\mu \Big|^{\kappa}$$

$$\leq \overline{\mathbb{E}}^{\square}_{\mathbf{h}\in(\mathbb{Z}^L)^s} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n,n'\in I_N} \prod_{m=1}^{\ell} T_{q_m(n;\mathbf{h})-q_\ell(n';\mathbf{h})} g_m(x;\mathbf{h}) \cdot$$

$$\prod_{m=1}^{\ell-1} T_{q_m(n';\mathbf{h})-q_\ell(n';\mathbf{h})} g_m(x;\mathbf{h}) \Big\|^{\kappa}_{L^2(\mu)}$$

$$= \overline{\mathbb{E}}^{\square}_{(\mathbf{h},\mathbf{h}')\in(\mathbb{Z}^{2L})^s} \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n,n'\in I_N} \prod_{m=1}^{2\ell-1} T_{q'_m((n,n');(\mathbf{h},\mathbf{h}'))} g'_m(x;(\mathbf{h},\mathbf{h}')) \Big\|^{\kappa}_{L^2(\mu)}$$

$$\leq S(A',\kappa),$$

where the last inequality holds because $(I_N \times I_N)_{N\in\mathbb{N}}$ is a Følner sequence of $\mathbb{Z}^L\times\mathbb{Z}^L$. On the other hand, if

$$q_i(n;\mathbf{h}) = \sum_{b,a_1,\dots,a_s\in\mathbb{N}^L} h_1^{a_1}\dots h_s^{a_s} n^b \cdot u_i(b;a_1,\dots,a_s)$$

for some $u_i(b;a_1,\dots,a_s)\in\mathbb{Q}^d$, then for $1\leq i\leq \ell-1$, we have

$$q'_{i+\ell}(n,n';\mathbf{h},\mathbf{h}') = \sum_{b,a_1,\dots,a_s\in\mathbb{N}^L} h_1^{a_1}\dots h_s^{a_s} n'^b \cdot (u_i(b;a_1,\dots,a_s) - u_\ell(b;a_1,\dots,a_s)),$$

and for $1\leq i\leq\ell$,

$$q'_i(n,n';\mathbf{h},\mathbf{h}') = \sum_{b,a_1,\dots,a_s\in\mathbb{N}^L} h_1^{a_1}\dots h_s^{a_s} (n^b\cdot u_i(b;a_1,\dots,a_s) - n'^b \cdot u_\ell(b;a_1,\dots,a_s)).$$

So for all $b,b',a_1,\dots,a_s,a'_1,\dots,a'_s\in\mathbb{N}^L$, similar to the argument in the proof of Proposition 6.2, we have

$$R_{\mathbf{q}}(b;a_1,\dots,a_s) = R_{\mathbf{q}'}((b,\mathbf{0});(a_1,\mathbf{0}),\dots,(a_s,\mathbf{0})) \sim R_{\mathbf{q}'}((\mathbf{0},b);(a_1,\mathbf{0}),\dots,(a_s,\mathbf{0}))$$

and $R_{q'}((b,b');(a_1,a'_1),\dots,(a_s,a'_s)) = \{\mathbf{0}\}$. This implies (29) and finishes the proof. □

We are now ready to prove Proposition 5.5 and close this section.

Proof of Proposition 5.5. Let A denote the PET-tuple $(L,0,k,(p_1,\dots,p_k),(f_1,\dots,f_k))$. Then for all $\kappa>0$,

$$S(A,\kappa) = \sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} \prod_{m=1}^{k} T_{p_m(n)} f_m \Big\|^{\kappa}_{L^2(\mu)}.$$

By the assumption, A is non-degenerate. We only prove (21) for f_1 as the other cases are identical.

We first assume that A is standard for f_1. By Theorem 4.2, there exist finitely many vdC-operations $\partial_{\rho_1}, \dots, \partial_{\rho_t}$ such that $A' = \partial_{\rho_t} \dots \partial_{\rho_1} A$ is a non-degenerate PET-tuple which is standard for f_1, and $\deg(A') = 1$. By Proposition 4.1, $S(A, 2^t) \leq C \cdot S(A', 1)$ for some $C > 0$ depending only on the polynomials $p_1, \dots, p_k$. We may assume that

$$S(A',1) = \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s \in \mathbb{Z}^L} \sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| \mathbb{E}_{n\in I_N} \prod_{m=1}^{\ell} T_{\mathbf{d}_m(h_1,\dots,h_s)\cdot n + r_m(h_1,\dots,h_s)} g_m(x; h_1, \dots, h_s) \Big\|_{L^2(\mu)}$$

for some $s, \ell \in \mathbb{N}^*$, functions $g_1, \dots, g_\ell \colon X \times (\mathbb{Z}^L)^s \to \mathbb{R}$, where $g_1(\cdot; h_1, \dots, h_s) = f_1$, such that each $g_m(\cdot; h_1, \dots, h_s)$ is an $L^\infty(\mu)$ function bounded by 1, and polynomials $\mathbf{d}_m \colon (\mathbb{Z}^L)^s \to (\mathbb{Z}^d)^L$ and $r_m \colon (\mathbb{Z}^L)^s \to \mathbb{Z}^d$, $1 \leq m \leq \ell$, where $\mathbf{d}_m, r_m$ take value in vectors with integer coordinates because the vdC-operations send integer-valued polynomials to integer-valued polynomials. Let $\mathbf{c}_{1,1} = -\mathbf{d}_1$ and $\mathbf{c}_{1,m} = \mathbf{d}_m - \mathbf{d}_1$ for $m \neq 1$. Since A' is non-degenerate, we have that $\mathbf{c}_{1,1}, \dots, \mathbf{c}_{1,s} \not\equiv \mathbf{0}$. By Proposition 6.1, we have that

$$\begin{aligned} S(A',1) &\leq C' \cdot \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s \in \mathbb{Z}^L} \|T_{r_1(h_1,\dots,h_s)} f_1\|_{\{G(\mathbf{c}_{1,i}(h_1,\dots,h_s))\}_{1\leq i\leq \ell}} \\ &= C' \cdot \overline{\mathbb{E}}^{\square}_{h_1,\dots,h_s \in \mathbb{Z}^L} \|f_1\|_{\{G(\mathbf{c}_{1,i}(h_1,\dots,h_s))\}_{1\leq i\leq \ell}} \end{aligned}$$

for some $C' > 0$ depending only on the polynomials $p_1, \dots, p_k$. Combining this with the fact that $S(A, 2^t) \leq C \cdot S(A', 1)$, we get (21).

Suppose that

$$\mathbf{c}_{1,m}(h_1, \dots, h_s) = \sum_{a_1,\dots,a_s \in \mathbb{N}^L} h_1^{a_1} \dots h_s^{a_s} \cdot \mathbf{u}_{1,m}(a_1, \dots, a_s), \quad \text{and}$$

$$\mathbf{d}_{1,m}(h_1, \dots, h_s) = \sum_{a_1,\dots,a_s \in \mathbb{N}^L} h_1^{a_1} \dots h_s^{a_s} \cdot \mathbf{v}_{1,m}(a_1, \dots, a_s)$$

for some $\mathbf{u}_{1,m}(a_1, \dots, a_s), \mathbf{v}_{1,m}(a_1, \dots, a_s) \in (\mathbb{Q}^d)^L$ with all but finitely many terms being zero for each m. Write $\mathbf{u}_{i,m}(a_1, \dots, a_s) = (u_{i,m,1}(a_1, \dots, a_s), \dots, u_{i,m,L}(a_1, \dots, a_s))$, $\mathbf{v}_{i,m}(a_1, \dots, a_s) = (v_{i,m,1}(a_1, \dots, a_s), \dots, v_{i,m,L}(a_1, \dots, a_s))$, and, for all $1 \leq r \leq \ell$, set

$$U_{1,r}(a_1, \dots, a_s) \coloneqq \{u_{1,m,r}(a_1, \dots, a_s) \in \mathbb{Q}^d \colon 1 \leq m \leq \ell\} \cup \{\mathbf{0}\}; \quad \text{and}$$

$$V_{1,r}(a_1, \dots, a_s) \coloneqq \{v_{1,m,r}(a_1, \dots, a_s) \in \mathbb{Q}^d \colon 1 \leq m \leq \ell\} \cup \{\mathbf{0}\}.$$

Since $A' = \partial_{w_t} \dots \partial_{w_1} A$, by repeatedly using Proposition 6.2, for all $a_1, \dots, a_s \in \mathbb{N}^L$ not all equal to $\mathbf{0}$ and every $1 \leq r \leq L$, there exists $v \in \mathbb{N}^L, v \neq \mathbf{0}$ such that $V_{1,r}(a_1, \dots, a_s) \lesssim R_v$. By the relation between $\mathbf{u}_{1,m}$ and $\mathbf{v}_{1,m}$, we get $U_{1,r}(a_1, \dots, a_s) \sim V_{1,r}(a_1, \dots, a_s)$ and so $U_{1,r}(a_1, \dots, a_s) \lesssim R_v$.

We now assume that $A = (L, 0, k, (p_1, \dots, p_k), (f_1, \dots, f_k))$ is not standard for f_1. Since A is semi-standard for f_1, by Proposition 6.3, there exists a PET-tuple $A' = (2L, 0, \ell, \mathbf{q}, \mathbf{g})$ which is non-degenerate and standard for f_1 such that $S(A, 2\kappa) \leq S(A', \kappa)$ for all $\kappa > 0$ and (29) holds. Working with the PET-tuple A' instead of A as before (and using (29)), we get the result. □

7. Proof of Proposition 5.6

This last section is dedicated to the proof of Proposition 5.6. We remark that it is in this proposition where the concatenation results (Theorem 2.10 and Corollary 2.11) are used.

Following the notation of Proposition 5.6, for every $\mathbf{h}=(h_1,\dots,h_s)\in(\mathbb{Z}^L)^s$ and $1\leq i\leq k$, we set

$$W_{i,\mathbf{h}} \coloneqq Z_{G(\mathbf{c}_{i,1}(\mathbf{h})),\dots,G(\mathbf{c}_{i,t_i}(\mathbf{h}))}(\mathbf{X}),$$

and for every subset $J\subseteq(\mathbb{Z}^L)^s$,

$$W_{i,J} \coloneqq \bigvee_{\mathbf{h}\in J} W_{i,\mathbf{h}}.$$

We have:

Lemma 7.1. *Let the notations be as in Proposition 5.6. If (23) holds for every $\mathbb{Z}^d$-system $(X,\mathcal{B},\mu,(T_g)_{g\in\mathbb{Z}^d})$ and every $f_1,\dots,f_k\in L^\infty(\mu)$, then for every $J_1,\dots,J_k\subseteq(\mathbb{Z}^L)^s$ of density* 1, *we have that*

$$\mathbb{E}_{n\in\mathbb{Z}^L} T_{p_1(n)}f_1\cdot\ldots\cdot T_{p_k(n)}f_k=0,\quad \textit{if}\quad \mathbb{E}(f_i|W_{i,J_i})=0 \textit{ for some } 1\leq i\leq k. \tag{30}$$

Proof. Suppose that $\mathbb{E}(f_i|W_{i,J_i})=0$ for some $1\leq i\leq k$. By definition, $\|f_i\|_{G(c_{i,1}(\mathbf{h})),\dots,G(c_{i,t_i}(\mathbf{h}))}=0$ for all $\mathbf{h}\in J_i$. Since J_i is of density 1, the conclusion follows from (23). ☐

Before proving Proposition 5.6, we continue with our main example (Example 1).

Third part of computations for Example 1: We are dealing with the $(T_1^{n^2+n},T_2^{n^2})$ case. Applying (26) to Lemma 7.1, we have that

$$\mathbb{E}_{n\in\mathbb{Z}} T_1^{n^2+n}f_1\cdot T_2^{n^2}f_2=0,\quad \text{if}\quad \mathbb{E}(f_i|W_{i,J_i})=0 \text{ for } i=1 \text{ or } 2, \tag{31}$$

for all $J_1,J_2\in\mathbb{Z}^3$ of density 1, where

$$W_{i,J}=\bigvee_{(h_1,h_2,h_3)\in J} W_{i,(h_1,h_2,h_3)}=\bigvee_{(h_1,h_2,h_3)\in J} Z_{G(\mathbf{c}_{i,1}(h_1,h_2,h_3)),\dots,G(\mathbf{c}_{i,7}(h_1,h_2,h_3))},\quad i=1,2,$$

where $\mathbf{c}_{i,j}\colon\mathbb{Z}^3\to\mathbb{Z}^2$ are the ones in the second part of computations for Example 1.

Recall that $e_1=(1,0)$, $e_2=(0,1)$, $e=e_1-e_2$. In this case, we have that $H_{1,1}=\mathbb{Z}e_1$, $H_{1,3}=H_{1,5}=H_{1,7}=\mathbb{Z}e$ and $H_{1,2}=H_{1,4}=H_{1,6}=\mathbb{Z}^2$. Moreover, $H_{2,1}=\mathbb{Z}e_2$, $H_{2,2}=H_{2,4}=H_{2,6}=\mathbb{Z}e$ and $H_{2,3}=H_{2,5}=H_{2,7}=\mathbb{Z}^2$.

Applying (26) to Lemma 7.1, we get

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}}\ \overline{\lim_{N\to\infty}}\left\|\mathbb{E}_{n\in I_N}\Big(T_1^{n^2+n}f_1\cdot T_2^{n^2}f_2-T_1^{n^2+n}\mathbb{E}(f_1|W_{1,\mathbb{Z}^3})\cdot T_2^{n^2}f_2\Big)\right\|_{L^2(\mu)}=0. \tag{32}$$

Fix $\varepsilon > 0$. Since $W_{1,\mathbb{Z}^3} = \bigvee_{N=1}^{\infty} W_{1,[-N,N]^3}$, by approximation, there exists a finite subset I of $\mathbb{Z}^3$ such that $\|\mathbb{E}(f_1|W_{1,\mathbb{Z}^3}) - \mathbb{E}(f_1|W_{1,I})\|_{L^1(\mu)} < \varepsilon^2/2$. Since $\|f_1\|_{L^\infty(\mu)}, \|f_2\|_{L^\infty(\mu)} \leq 1$, for all $n \in \mathbb{Z}$,

$$\Big\|\Big(T_1^{n^2+n}\mathbb{E}(f_1|W_{1,\mathbb{Z}^3}) \cdot T_2^{n^2} f_2 - T_1^{n^2+n}\mathbb{E}(f_1|W_{1,I}) \cdot T_2^{n^2} f_2\Big)\Big\|_{L^2(\mu)}^2$$

$$= \int_X \Big(T_1^{n^2+n}\mathbb{E}(f_1|W_{1,\mathbb{Z}^3}) \cdot T_2^{n^2} f_2 - T_1^{n^2+n}\mathbb{E}(f_1|W_{1,I}) \cdot T_2^{n^2} f_2\Big)^2 \, d\mu$$

$$\leq \int_X 2\Big|T_1^{n^2+n}\mathbb{E}(f_1|W_{1,\mathbb{Z}^3}) - T_1^{n^2+n}\mathbb{E}(f_1|W_{1,I})\Big| \, d\mu = \int_X 2\Big|\mathbb{E}(f_1|W_{1,\mathbb{Z}^3}) - \mathbb{E}(f_1|W_{1,I})\Big| \, d\mu < \varepsilon^2.$$

So,

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N}\Big(T_1^{n^2+n}\mathbb{E}(f_1|W_{1,\mathbb{Z}^3}) \cdot T_2^{n^2} f_2 - T_1^{n^2+n}\mathbb{E}(f_1|W_{1,I}) \cdot T_2^{n^2} f_2\Big)\Big\|_{L^2(\mu)} < \varepsilon. \tag{33}$$

Note that $W_{1,I}$ is contained in the $(7|I|)$-step factor

$$W_1' := Z_{(G(\mathbf{c}_{1,1}(h_1,h_2,h_3)),\dots,G(\mathbf{c}_{1,7}(h_1,h_2,h_3)))_{(h_1,h_2,h_3)\in I}}.$$

We say that $(h_1', h_2', h_3') \in \mathbb{Z}^3$ is *good* if for any $(h_1, h_2, h_3) \in I$, any

$$g \in \{-2h_1e_1, 2h_2e - 2h_1e_1, 2h_2e, 2h_3e - 2h_1e_1, 2h_3e, 2(h_2+h_3)e - 2h_1e_1, 2(h_2+h_3)e\}$$

(*i.e.*, g is the generator of one of $G(\mathbf{c}_{1,1}(h_1,h_2,h_3)), \dots, G(\mathbf{c}_{1,7}(h_1,h_2,h_3)))$ and any action

$$g' \in \{-2h_1'e_1, 2h_2'e - 2h_1'e_1, 2h_2'e, 2h_3'e - 2h_1'e_1, 2h_3'e, 2(h_2'+h_3')e - 2h_1'e_1, 2(h_2'+h_3')e\},$$

(*i.e.*, g' is the generator of one of $G(\mathbf{c}_{1,1}(h_1',h_2',h_3')), \dots, G(\mathbf{c}_{1,7}(h_1',h_2',h_3')))$ the set

$$H := \text{span}_{\mathbb{Q}}\{g, g'\} \cap \mathbb{Z}^2$$

satisfies the following:

$$\begin{cases} H = \mathbb{Z}e_1 & \text{, if } g = -2h_1e_1, g' = -2h_1'e_1 \\ H = \mathbb{Z}e & \text{, if } g \in \{2h_2e, 2h_3e, 2(h_2+h_3)e\}, g' \in \{2h_2'e, 2h_3'e, 2(h_2'+h_3')e\} \\ H = \mathbb{Z}^2 & \text{, otherwise} \end{cases}.$$

Let J be the set of all good tuples. Since I is finite, it is not hard to show that J is of density 1 (see also the claim in the proof of Proposition 5.6). Again, applying (26) to Lemma 7.1,

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N}\Big(T_1^{n^2+n}\mathbb{E}(f_1|W_{1,I}) \cdot T_2^{n^2} f_2 - T_1^{n^2+n}\mathbb{E}(f_1|W_{1,J} \cap W_{1,I}) \cdot T_2^{n^2} f_2\Big)\Big\|_{L^2(\mu)} = 0. \tag{34}$$

So (32), (33) and (34) implies that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}} \\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N}\Big(T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2 - T_1^{n^2+n}\mathbb{E}(f_1|W_{1,J} \cap W_{1,I}) \cdot T_2^{n^2} f_2\Big)\Big\|_{L^2(\mu)} < \varepsilon. \tag{35}$$

By the definition of good tuples and Corollary 2.11, we have that

$$W_{1,J} \cap W_{1,I} \subseteq \bigvee_{(h_1',h_2',h_3')\in J} W_1' \cap W_{1,(h_1',h_2',h_3')} = \bigvee_{(h_1',h_2',h_3')\in J} Z_{(\mathbb{Z}e_1)^{\times|I|},(\mathbb{Z}e)^{\times 9|I|},(\mathbb{Z}^2)^{\times 39|I|}} \subseteq Z_{e^{\times\infty},e_1^{\times\infty}}.$$

So (35) implies that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N} T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2\Big\|_{L^2(\mu)} < \varepsilon, \quad \text{if} \quad \mathbb{E}(f_1|Z_{e_1^{\times\infty}, e^{\times\infty}}) = 0.$$

Since $\varepsilon > 0$ is arbitrary,

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N} T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2\Big\|_{L^2(\mu)} = 0, \quad \text{if} \quad \mathbb{E}(f_1|Z_{e_1^{\times\infty}, e^{\times\infty}}) = 0.$$

Working analogously for the $T_2^{n^2} f_2$ term, we eventually get that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\|\mathbb{E}_{n\in I_N} T_1^{n^2+n} f_1 \cdot T_2^{n^2} f_2\Big\|_{L^2(\mu)} = 0 \ \text{ if } \ \mathbb{E}(f_i|Z_{e_i^{\times\infty}, e^{\times\infty}}) = 0 \ \text{ for } i = 1 \text{ or } 2. \tag{36}$$

We remark that (36) is a stronger version of (14) (*i.e.*, continuation of Example 1).

Remark. As mentioned before, the characteristic factors described in Theorem 5.1 are not the optimal ones in general, but they are good enough for the purposes of our study.

We briefly explain the idea on proving Proposition 5.6. Under the assumptions of Proposition 5.6, Lemma 7.1 says that one can assume that f_1 is measurable with respect to the factor W_{1,J_1}. However, thanks to the freedom of the choices of J_1, we can use Lemma 7.1 to repeatedly choose different subsets $J_{1,1}, \dots, J_{1,r}$, for some $r \in \mathbb{N}^*$, and assume that f_1 is measurable with respect to the factor $W_{1,J_{1,1}} \cap W_{1,J_{1,2}} \cap \dots \cap W_{1,J_{1,r}}$. We then employ the concatenation theorems to estimate the intersection of $W_{1,J_{1,j}}$, and find a smaller factor characterizing the multiple average we aim to study.

Proof of Proposition 5.6. Let $(X, \mathcal{B}, \mu, (T_g)_{g\in\mathbb{Z}^d})$ be a $\mathbb{Z}^d$-system, $f_1, \dots, f_k \in L^\infty(\mu)$ and $s, t_1, \dots, t_k, \mathbf{c}_{i,m}$, $1 \le i \le k$, $1 \le m \le t_i$ be as in the statement. By Lemma 2.13,

$$H_{i,m} = \mathrm{span}_{\mathbb{Q}}\{G(\mathbf{c}_{i,m}(h_1, \dots, h_s))\colon h_1, \dots, h_s \in \mathbb{Z}^L\} \cap \mathbb{Z}^d.$$

To show (24), it suffices to show that if $\mathbb{E}(f_i|Z_{(H_{i,1})^{\times\infty}, \dots, (H_{i,t_i})^{\times\infty}}) = 0$ for some $1 \le i \le k$, then the left hand side of (24) equals to 0. We assume without loss of generality that $i = 1$.

For every $r \in \mathbb{N}$, every finite subset $I \subseteq \mathbb{Z}^L$, and every tuple $(J_1, \dots, J_r)$, where $J_i \subseteq (\mathbb{Z}^L)^s$, $1 \le i \le r$, denote

$$A_I(J_1, \dots, J_r) \coloneqq \mathbb{E}_{n\in I} T_{p_1(n)} \mathbb{E}(f_1|W_{1,J_1} \cap \dots \cap W_{1,J_r}) \cdot T_{p_2(n)} f_2 \cdot \ldots \cdot T_{p_k(n)} f_k,$$

and in the degenerated case, set

$$A_I(\emptyset) \coloneqq \mathbb{E}_{n\in I} T_{p_1(n)} f_1 \cdot T_{p_2(n)} f_2 \cdot \ldots \cdot T_{p_k(n)} f_k.$$

We say that a tuple $(J_1, \dots, J_r)$ of subsets of $(\mathbb{Z}^L)^s$ is *admissible* if for every $\mathbf{h}_u \in J_u, 1 \le u \le r$ and every $1 \le m \le t_1$, denoting

$$G_K \coloneqq \mathrm{span}_{\mathbb{Q}}\{G(\mathbf{c}_{1,m}(\mathbf{h}_u))\colon u \in K\} \cap \mathbb{Z}^d \tag{37}$$

for all $K \subseteq \{1, \dots, r\}$, the following holds: for all $\emptyset \neq K' \subsetneq K \subseteq \{1, \dots, r\}$ such that $\max\{x \in K'\} < \min\{x \in K \backslash K'\}$, either $G_{K'} \subsetneq G_K$ or $G_{K'} = H_{1,m}$.[36]

[36] We think of this as a notion of having "full rank".

Fix $\varepsilon > 0$. By (23), we have that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| A_{I_N}(\emptyset) - A_{I_N}((\mathbb{Z}^L)^s) \right\|_{L^2(\mu)} = 0.$$

By an approximation argument similar to the one obtaining (33), there exists a finite subset $J_1' \subseteq (\mathbb{Z}^L)^s$ such that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| A_{I_N}((\mathbb{Z}^L)^s) - A_{I_N}(J_1') \right\|_{L^2(\mu)} < \varepsilon,$$

and so

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| A_{I_N}(\emptyset) - A_{I_N}(J_1') \right\|_{L^2(\mu)} < \varepsilon.$$

Suppose that for some $r \geq 1$, we have constructed finite subsets $J_1', \dots, J_r' \subseteq (\mathbb{Z}^L)^s$ such that:

(i) $\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| A_{I_N}(\emptyset) - A_{I_N}(J_1', \dots, J_r') \right\|_{L^2(\mu)} < r\varepsilon$; and

(ii) $(J_1', \dots, J_r')$ is admissible.

We now construct J_{r+1}'. We first claim that there exists $J_{r+1} \subseteq (\mathbb{Z}^L)^s$ of density 1 such that $(J_1', \dots, J_r', J_{r+1})$ is admissible. For every $\mathbf{h}_u \in J_u'$, $1 \leq u \leq r$, $1 \leq m \leq t_1$ and nonempty subset $K \subseteq \{1, \dots, r\}$, let

$$Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} := \text{span}_{\mathbb{Q}}\{G(\mathbf{c}_{1,m}(\mathbf{h}_u)) \colon u \in K\} \cap \mathbb{Z}^d.$$

If $Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} = H_{1,m}$, we let $V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} = (\mathbb{Z}^L)^s$, otherwise $V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K}$ denotes the set of $\mathbf{h} = (h_1, \dots, h_s) \in (\mathbb{Z}^L)^s$ such that $G(\mathbf{c}_{1,m}(\mathbf{h}))$ is not contained in $Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K}$. Let

$$J_{r+1} := \bigcap_{\mathbf{h}_u\in I_u,\ 1\leq u\leq r,\ 1\leq m\leq t_1,\ K\subseteq\{1,\dots,r\}} V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K}.$$

To show that $(J_1', \dots, J_r', J_{r+1})$ is admissible, fix $\mathbf{h}_i \in J_i'$, $1 \leq i \leq r$, $\mathbf{h}_{r+1} \in J_{r+1}$, $1 \leq m \leq t_1$, and let G_K be defined as in (37) for all $K \subseteq \{1, \dots, r+1\}$. Let $\emptyset \neq K' \subsetneq K \subseteq \{1, \dots, r+1\}$ such that $\max\{x \in K'\} < \min\{x \in K \backslash K'\}$. We have the following three possible cases for $r+1$:

Case (i): $r + 1 \notin K$**.** Then $r + 1 \notin K'$ and so $\emptyset \neq K' \subsetneq K \subseteq \{1, \dots, r\}$. Since $(I_1, \dots, I_r)$ is admissible, either $G_{K'} \subsetneq G_K$ or $G_{K'} = H_{1,m}$.

Case (ii): $r+1 \in K'$. This contradicts the assumption that $\max\{x \in K'\} < \min\{x \in K\backslash K'\}$. So this case is not possible.

Case (iii): $r + 1 \in K$ but $r + 1 \notin K'$. Then $K' \subseteq \{1, \dots, r\}$ and so $J_{r+1} \subseteq V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K'}$. If $G_{K'} \neq H_{1,m}$, then since $\mathbf{h}_{r+1} \in J_{r+1} \subseteq V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K'}$, the subgroup $G(\mathbf{c}_{1,m}(\mathbf{h}_{r+1}))$ (which is contained in G_K since $r + 1 \in K$) is not contained in $Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K'} = G_{K'}$. This implies that $G_K \neq G_{K'}$.

In conclusion, we have that $(J_1', \dots, J_r', J_{r+1})$ is admissible. The second part of the claim is that J_{r+1} is of density 1. Since $J_1', \dots, J_r'$ are finite sets, it suffices to show that every $V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K}$ is of density 1. If $Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} = H_{1,m}$, $V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} = (\mathbb{Z}^L)^s$ and we are done. Now assume that $Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} \neq H_{1,m}$. By Lemma 2.12, the set

$$V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} = \{\mathbf{h} \in (\mathbb{Z}^L)^s \colon G(\mathbf{c}_{1,m}(\mathbf{h})) \nsubseteq Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K}\}$$

is either of density 1, or is empty and

$$Q_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K} = \mathrm{span}_{\mathbb{Q}}\{G(\mathbf{c}_{i,m}(h_1,\dots,h_s))\colon h_1,\dots,h_s \in \mathbb{Z}^L\} \cap \mathbb{Z}^d = H_{1,m}.$$

By our assumption, $V_{m;\mathbf{h}_1,\dots,\mathbf{h}_r;K}$ is of density 1. This finishes the proof of the claim.

By Lemma 7.1, $A(J'_1,\dots,J'_r) = A(J'_1,\dots,J'_r,J_{r+1})$. By an approximation argument, there exists a finite subset $J'_{r+1} \subseteq J_{r+1}$ such that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| A_{I_N}(J'_1,\dots,J'_r,J_{r+1}) - A_{I_N}(J'_1,\dots,J'_r,J'_{r+1}) \Big\|_{L^2(\mu)} < \varepsilon.$$

By induction hypothesis,

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| A_{I_N}(\emptyset) - A_{I_N}(J'_1,\dots,J'_r,J'_{r+1}) \Big\|_{L^2(\mu)} < (r+1)\varepsilon.$$

So (i) holds for $r+1$. Since $(J'_1,\dots,J'_r,J_{r+1})$ is admissible, so is $(J'_1,\dots,J'_r,J'_{r+1})$, hence (ii) holds for $r+1$. In conclusion, there exist a tuple $(J'_1,\dots,J'_{dt_1})$ of finite subsets of $(\mathbb{Z}^L)^s$ such that

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \Big\| A_{I_N}(\emptyset) - A_{I_N}(J'_1,\dots,J'_{dt_1}) \Big\|_{L^2(\mu)} < dt_1\varepsilon$$

and $(J'_1,\dots,J'_{dt_1})$ is admissible. Note that

$$\begin{aligned} W_{1,J'_1} \cap \dots \cap W_{1,J'_{dt_1}} &= \bigcap_{u=1}^{dt_1} \bigvee_{\mathbf{h}_u \in J'_u} W_{1,\mathbf{h}_u} \\ &= \bigcap_{u=1}^{dt_1} \bigvee_{\mathbf{h}_u \in J'_u} Z_{G(\mathbf{c}_{1,1}(\mathbf{h}_u)),\dots,G(\mathbf{c}_{1,t_1}(\mathbf{h}_u))} \\ &\subseteq \bigcap_{u=1}^{dt_1} Z_{\{G(\mathbf{c}_{1,m}(\mathbf{h}_u))\}_{1\le m\le t_1,\mathbf{h}_u\in J'_u}}, \end{aligned}$$

where we used Lemma 2.4 (vii) in the last inclusion. For each $1 \le u \le dt_1$, pick some $1 \le m_u \le t_1$ and $\mathbf{h}_u \in J'_u$. Consider the set

$$P \coloneqq \mathrm{span}_{\mathbb{Q}}\{G(\mathbf{c}_{1,m_u}(\mathbf{h}_u))\colon 1 \le u \le dt_1\} \cap \mathbb{Z}^d.$$

By the pigeon-hole principle, there exist $1 \le m \le t_1$ and $1 \le u_1 < \dots < u_d \le dt_1$ such that $m_{u_1} = \dots = m_{u_d} = m$. For all $1 \le i \le d$, let $K_i = \{u_1,\dots,u_i\} \subseteq \{1,\dots,dt_1\}$ and

$$P_i \coloneqq \mathrm{span}_{\mathbb{Q}}\{G(\mathbf{c}_{1,m_u}(\mathbf{h}_u))\colon u \in K_i\} \cap \mathbb{Z}^d.$$

Since $(J'_1,\dots,J'_{dt_1})$ is admissible, for all $1 \le i \le d-1$, either $P_i = H_{1,m}$ or the dimension of P_{i+1} is higher than that of P_i. Since the dimension of P_i can not exceed d, we must have that P_i contains $H_{1,m}$ for some $1 \le i \le d$. As $P_i \subseteq P$, we have that P also contains $H_{1,m}$. By Corollary 2.11,

$$W_{1,J'_1} \cap \dots \cap W_{1,J'_{dt_1}} \subseteq \bigcap_{u=1}^{dt_1} Z_{\{G(\mathbf{c}_{1,m}(\mathbf{h}_u))\}_{1\le m\le t_1,\mathbf{h}_u\in J'_u}} \subseteq Z_{H_{1,1}^{\times\infty},\dots,H_{1,t_1}^{\times\infty}}.$$

Since $\mathbb{E}(f_1|Z_{(H_{1,1})^{\times\infty},\ldots,(H_{1,t_1})^{\times\infty}})=0$, $A(J_1',\ldots,J_{dt_1}')=0$ and so

$$\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| A_{I_N}(\emptyset)\right\|_{L^2(\mu)} < dt_1\varepsilon.$$

Since ε is chosen arbitrary, the left hand side of (24) is equal to $\sup_{\substack{(I_N)_{N\in\mathbb{N}}\\ \text{Følner seq.}}} \overline{\lim_{N\to\infty}} \left\| A_{I_N}(\emptyset)\right\|_{L^2(\mu)} = 0$, which finishes the proof. □

References

[1] T. Austin. On the norm convergence of nonconventional ergodic averages. *Ergodic Theory Dynam. Systems* **30** (2010), no. 2, 321–338.

[2] T. Austin. Pleasant extensions retaining algebraic structure, I. *J. Anal. Math.* **125** (2015), 1–36.

[3] T. Austin. Pleasant extensions retaining algebraic structure, II. *J. Anal. Math.* **126** (2015), 1–111.

[4] D. Berend and V. Bergelson. Jointly ergodic measure-preserving transformations. *Israel J. Math.* **49** (1984), no. 4, 307–314.

[5] V. Bergelson. Ergodic Ramsey theory–an update. Ergodic theory of $\mathbf{Z}^d$ actions (Warwick, 1993–1994), 1–61, London Math. Soc. Lecture Note Ser., 228, Cambridge Univ. Press, Cambridge, 1996.

[6] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems* **7** (1987), no. 3, 337–349.

[7] V. Bergelson and A. Leibman. Cubic averages and large intersections. *Recent trends in ergodic theory and dynamical systems*, 5–19, Contemp. Math., 631, Amer. Math. Soc., Providence, RI, 2015.

[8] V. Bergelson, A. Leibman and Y. Son. Joint ergodicity along generalized linear functions. *Ergodic Theory Dynam. Systems* **36** (2016), no. 7, 2044–2075.

[9] Q. Chu, N. Frantzikinakis and B. Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. *Proc. of the London Math. Society*, (3), **102** (2011), 801–842.

[10] M. Einsiedler and T. Ward. *Ergodic theory with a view towards number theory.* Springer-Verlag, London, **259** (2011).

[11] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.* **31** (1977), 204–256.

[12] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, Princeton, 1981.

[13] N. Frantzikinakis. A multidimensional Szemerédi theorem for Hardy sequences of different growth. *Tran. of the A. M. S.* **367**, no. 8, (2015), 5653–5692.

[14] N. Frantzikinakis and B. Kra. Polynomial averages converge to the product of integrals. *Israel J. Math.* **148** (2005), 267–276.

[15] B. Host. Ergodic seminorms for commuting transformations and applications. *Studia Math.*, **195** (2009), no.1, 31–49.

[16] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math.* (2) **161** (2005), no. 1, 397–488.

[17] M. Johnson. Convergence of polynomial ergodic averages of several variables for some commuting transformations. *Illinois J. Math.* **53** (2009), no. 3, 865–882.

[18] D. Karageorgos and A. Koutsogiannis. Integer part independent polynomial averages and applications along primes, to appear in *Studia Mathematica*.

[19] A. Koutsogiannis. Integer part polynomial correlation sequences. *Ergodic Theory Dynam. Systems* **38** (2018), no. 4, 1525–1542.

[20] A. Leibman. Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, 215–225.

[21] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Israel J. Math.* **146** (2005), 303–315.

[22] W. Sun. Weak ergodic averages over dilated curves. Submitted for publication. arXiv: 1809.06916.

[23] T. Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dynam. Systems* **28** (2008), no. 2, 657–688.
[24] T. Tao and T. Ziegler. Concatenation Theorems for anti-Gowers-uniform functions and Host-Kra characteristic factors. *Discrete Anal.* 2016, Paper No. 13.
[25] M. Walsh. Norm convergence of nilpotent ergodic averages. *Ann. of Math.* (2) **175** (2012), no. 3, 1667–1688.
[26] T. Ziegler. Nilfactors of $\mathbb{R}^m$-actions and configurations in sets of positive upper density in $\mathbb{R}^m$. *J. Anal. Math.*, **99** (2006), 249–266.
[27] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20** (2007), no. 1, 53–97.

(Sebastián Donoso) Universidad de O'Higgins, Instituto de Ciencias de la Ingeniería, Av. Libertador Bernardo O'Higgins 611, Rancagua, 2841959, Chile
E-mail address: sebastian.donoso@uoh.cl

(Andreas Koutsogiannis) The Ohio State University, Department of mathematics, 231 West 18th Avenue, Columbus, OH 43210-1174, USA
E-mail address: koutsogiannis.1@osu.edu

(Wenbo Sun) The Ohio State University, Department of mathematics, 231 West 18th Avenue, Columbus, OH 43210-1174, USA
E-mail address: sun.1991@osu.edu